Cognition 274 (2026) 106598

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full Length Article

Curiosity and evidence quality affect information seeking and eye movements during reading

Jaana Simola [a,b,*], Hanna Julku [a], Caitlin Dawson [a], Sebastian Österman [a], Kaisa Elovaara [a,c], Tuomo Häikiö [c]

[a] Department of Education, University of Helsinki, Siltavuorenpenger 3A (P.O. Box 9), 00014 University of Helsinki, Finland
[b] Cognitive Brain Research Unit (CBRU), University of Helsinki, Siltavuorenpenger 5A (P.O. Box 9) / Haartmaninkatu 3 (P.O. Box 21), 00014 University of Helsinki, Finland
[c] Department of Psychology and Speech-Language Pathology, University of Turku, Assistentinkatu 7, 20500 Turku, Finland

ARTICLE INFO

Keywords:
curiosity
eye-movements
information seeking
reading
exploration-exploitation

ABSTRACT

Curiosity is critical for learning and selective information processing, but it is unclear how it affects eye-movement patterns during reading. We tracked participants' (*n*=56) eye movements in a task with 18 health-related questions. Each question was followed by up to 15 statements from three quality categories: scientific relevant (SR), scientific irrelevant (SI), and non-scientific relevant (NR). The statements were concealed in three boxes according to evidence quality. Participants could click on any box to reveal a new statement until they had read all the statements or were ready to answer the question. After reading a statement, participants rated their state curiosity and how useful the statement was for answering the question. The SR statements were rated highest on usefulness and were read most frequently, suggesting that participants were able to detect that these statements had the highest evidence quality. For a more comprehensive view on curiosity, we also measured participants' trait curiosity with a questionnaire. We found that state curiosity correlated negatively with the thrill-seeking dimension of trait curiosity. Eye movements during statement reading were analyzed at the word-level, controlling for the effects of word length and frequency. Higher state curiosity was associated with lower word-skipping rates and longer total fixation duration on words, irrespective of evidence quality. However, the increase of total fixation duration with state curiosity was steeper for low quality (SI and NR) than high quality (SR) statements. Together, these results suggest that curiosity is related to a careful reading strategy, particularly for low-quality statements.

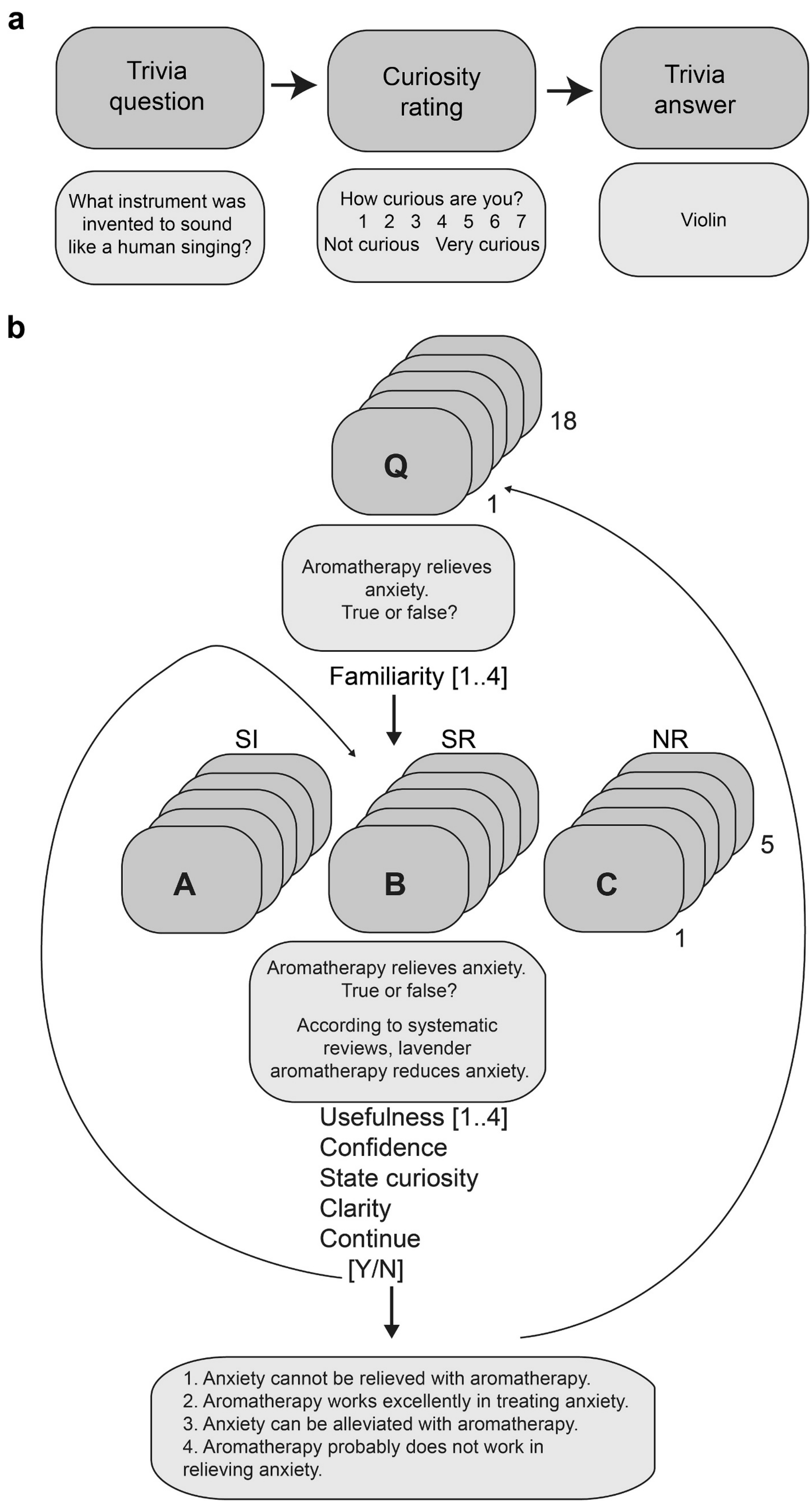
1. Introduction

Curiosity motivates humans to learn and is therefore foundational to our cognition. Curiosity is defined as motivation for exploration and novelty seeking (Berlyne, 1954b), and a specific type of curiosity, *epistemic curiosity,* refers to our desire for knowledge (Litman, 2008; Litman & Spielberger, 2003; Loewenstein, 1994; Markey & Loewenstein, 2014). Epistemic curiosity has been studied with trivia question paradigms, where on each trial, participants are presented with a question. Before viewing the answer, they are asked to rate how curious they are to learn the correct answer (Fig. 1a). Curiosity ratings are then correlated with behavioral performance or neural activity around the onset of the questions, or when anticipating or viewing the answers. These tasks provide a powerful way to investigate neural (Gruber, Gelman, & Ranganath, 2014; Kang et al., 2009; Ligneul, Mermillod, & Morisseau, 2018), behavioral (Marvin & Shohamy, 2016), and eye movement (Baranes, Oudeyer, & Gottlieb, 2015) correlates of momentary experiences of curiosity. However, the ecological validity of these tasks has been questioned as they provide a clear and explicit knowledge gap that is closed when the answer is revealed (see Schumacher et al., 2026).

In real-world situations, curiosity may arise more spontaneously and require sustained focus on a topic over longer periods of time, such as reading an article, conducting an online search, or taking a course. Reading tasks allow investigations of curiosity in more naturalistic settings (Dawson et al., 2024; Lydon-Staley, Zhou, Blevins, Zurn, & Bassett, 2021; Schumacher et al., 2026), and compared to self-reported curiosity, reading times provide an unobtrusive and behavioral indicator of

* Corresponding author at: Department of Education, University of Helsinki, Siltavuorenpenger 3A (P.O. Box 9), 00014 University of Helsinki, Finland.
E-mail address: jaana.simola@helsinki.fi (J. Simola).

https://doi.org/10.1016/j.cognition.2026.106598
Received 26 June 2025; Received in revised form 18 May 2026; Accepted 25 May 2026
Available online 4 June 2026



(caption on next page)

Fig. 1. *Schematics of the trial structure.* a) In trivia question paradigms, participants are presented with a trivia question. After reading each question, participants rate their curiosity about its answer and are then presented with the correct answer (e.g., Gruber et al., 2014; Kang et al., 2009). b) In the 3-armed bandit paradigm used in the present study, participants were presented with 18 health-related questions (Q). After reading a question, participants rated their familiarity with the topic of the question from 1 (low) to 4 (high). Next, three boxes (A, B, and C) were presented on the screen, and participants were asked to choose a box and read a randomly selected statement from that box. The boxes contained statements from three different evidence quality categories: scientific irrelevant (SI), scientific relevant (SR), and non-scientific relevant (NR). The evidence quality of the boxes remained the same for the duration of one question, but the quality categories were shuffled between questions. To learn about the contents of the boxes for each question, participants had to acquire knowledge over the rounds of reading the statements. The task allowed us to quantify switches between the boxes as *exploration* and choosing a statement from the same box than on the previous round as *exploitation*. Note that the questions were presented alongside the statements to avoid participants having to memorize the questions, but the eye movement analysis focused solely on the statements. After reading a statement, participants rated it on its usefulness, their confidence about the usefulness rating, their state curiosity, and its clarity from 1 to 4. They could read as many statements as they wished or until the statement selection was exhausted (as indicated by the arrow on the left) before moving on to answer the multiple-choice question. After answering the question, a round with a new question started (as indicated by the arrow on the right).

curiosity. However, less is known about how curiosity affects the reading process as it unfolds. The present study investigated how self-reported curiosity is related to information seeking behavior and the reading process as reflected in participants' eye movement patterns, while they were reading statements about health-related questions.

Curiosity is influenced by situational and dispositional processes, reflecting the division of curiosity between state and trait curiosity. *State curiosity* refers to a temporary experience of curiosity and a cognitive state specific to a given situation (Gruber & Ranganath, 2019; Loewenstein, 1994; Markey & Loewenstein, 2014). According to this account, curiosity is a motivated state of information seeking that is induced by certain topics or experiences. The notion of state curiosity is contrasted with *trait curiosity*, which refers to a personality disposition toward experiencing curiosity. This line of research proposes that curiosity is a multi-dimensional and stable trait that varies across individuals (Eschmann, Pereira, Valji, Dehmelt, & Gruber, 2023; Kahan, Landrum, Carpenter, Helft, & Hall Jamieson, 2017; Kashdan, Disabato, Goodman, & McKnight, 2020; Litman & Spielberger, 2003). For example, according to the Five-Dimensional Curiosity Scale Revised (5DCR) (Kashdan et al., 2020), trait curiosity is comprised of five dimensions, two of which are related to epistemic curiosity, reflecting the emotional valence of facing an information gap (joyous exploration, deprivation sensitivity), two reflect an individual's approach and response to novelty (thrill seeking, stress tolerance), and the final dimension covers both overt and covert curiosity about other people.

According to appraisal theories of emotion, individual differences in emotions are driven by individual differences in patterns of cognitive appraisals (Scherer, 2001). Appraisals are evaluations of how events relate to one's goals, values, knowledge, and abilities. Appraisal theories suggest a link between state and trait measures of emotion. In the context of curiosity, frequent experiences of state-level curiosity may develop into a more stable form of interest (i.e., trait curiosity, Fayn, Silvia, Dejonckheere, Verdonck, & Kuppens, 2019; Silvia, 2008). That is, curious people develop tendencies to feel curious because they often make appraisals that cause interest. Similar results have been obtained for other personality traits (e.g., the "Big Five"), such that each person's scores for personality traits correspond with their mean in a density distribution of situation-specific states related to the personality traits (Fleeson & Gallagher, 2009; Fleeson & Jayawickreme, 2015). In line with these accounts, prior research indicates that state curiosity correlates with trait curiosity (Litman, Hutchins, & Russon, 2005). As trait curiosity describes a tendency to experience state curiosity, to obtain a more comprehensive view on curiosity, we investigated how trait-level curiosity (as measured with the 5DCR) was related to participants' information seeking behavior and their state curiosity during the reading task.

Epistemic curiosity is based on metacognition about one's current knowledge and can be evoked by an information gap, a discrepancy between what a person knows and what they want to know (Loewenstein, 1994). It can also be prompted by cognitive incongruity, such as contradictions between expectations and new information. Deviations from expectations elicit surprises (Vogl, Pekrun, Murayama, & Loderer, 2020) that activate curiosity (Berlyne, 1954a; Loewenstein, 1994). Silvia (2010) has proposed that curiosity arises when discrepant information is perceived as novel and complex, but comprehensible. Curiosity thus relies on prior knowledge. Studies have shown that familiar items evoke higher curiosity than less familiar items (Berlyne, 1954a) and that metacognition is related to curiosity in an inverted-U shaped function (Kang et al., 2009; Kidd & Hayden, 2015; Metcalfe, Schwartz, & Eich, 2020): Curiosity peaks if an individual has intermediate familiarity or confidence about their knowledge but declines if familiarity or confidence is too high or low (i.e., those who know very little have not had their curiosity piqued and those who know a lot are satiated and not curious to know more). Similarly, studies among 7- to 8-month-old infants have shown that predictability in both visual (Kidd, Piantadosi, & Aslin, 2012) and auditory (Kidd, Piantadosi, & Aslin, 2014) stimulus stream influences allocation of attention. That is, infants show a greater likelihood of allocating attention to stimuli that were moderately novel and surprising (complex) in relation to previously presented stimuli. Like curiosity, infants' attention peaked for stimuli with intermediate levels of complexity, i.e., for stimuli falling in a "Goldilocks" range, exhibiting an inverted U-shaped pattern (Kidd et al., 2012; Kidd et al., 2014).

Moreover, memory is enhanced for stimuli eliciting curiosity (Gruber et al., 2014; Gruber & Ranganath, 2019; Kang et al., 2009; Ligneul et al., 2018; Marvin & Shohamy, 2016) and it is enhanced even for task-unrelated materials when participants were exposed to them during high-curiosity states (Gruber et al., 2014). Together, these results highlight the importance of curiosity in memory and learning. To investigate whether curiosity was related to memory in our task, we asked participants to respond to the health-related questions in an online survey approximately two weeks after the lab experiment and examined whether state curiosity and information seeking during the lab task affected these responses.

1.1. Curiosity, attention, and eye movements

According to the information-gap theory, curiosity arises when attention is focused on a gap in one's knowledge (Loewenstein, 1994) and it declines when attention is drawn away from the information gap (Markey & Loewenstein, 2014). Studies have also proposed that curiosity-based attention increases when people are in a metacognitive state where they feel that they are on the verge of knowing or understanding (Metcalfe et al., 2020), and that optimal learning occurs when the learners focus their attention on items that they perceive to be in the 'almost known' zone. Enhanced curiosity in this zone is proposed to be linked with decreased mind wandering (see Metcalfe et al., 2020). Mind wandering is a state characterized by the decoupling of attention from the current task to internal, self-generated, and task-unrelated thoughts and feelings (Smallwood & Schooler, 2015) and is considered an antithesis of curiosity (Metcalfe et al., 2020).

In complex and realistic choices involving active sampling of the environment, curiosity is used to assign value to specific types of information (Gottlieb & Oudeyer, 2018; Gottlieb, Oudeyer, Lopes, & Baranes, 2013). Eye movements are natural indicators of active information sampling. Converging neurophysiological data indicate that curiosity can bias attention and eye movements by weighting sensory cues based on intrinsic motivations aimed at reducing uncertainty (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018). An example of an active and instrumental sampling (for extrinsic reward purposes) is a situation where a driver approaches an intersection and attends to the traffic light to reduce uncertainty about whether to accelerate or to brake.

The few studies that have investigated the effects of curiosity on eye movements have found both trait and state curiosity to affect eye movements. For example, participants with higher perceptual curiosity (i.e., curiosity towards sensory stimuli and new perceptual experiences), engaged in more widespread saccadic exploration of visual scenes including buildings, interiors and landscapes, whereas the non-sensory aspects of trait curiosity did not correlate with eye movement patterns (Risko, Anderson, Lanthier, & Kingstone, 2012). Moreover, eye movements collected with a head-mounted eye-tracking device, while participants were running an errand on a university campus, were found to predict perceptual curiosity (Hoppe, Loetscher, Morey, & Bulling, 2018). Even more limited research has examined the effect of epistemic curiosity on eye movements. Gross, Araujo, Zedelius, and Schooler (2019) investigated the effects of creativity, schizotypy, and perceptual as well as epistemic curiosity on eye movements during free viewing of images, including natural indoor and outdoor scenes as well as non-representational artworks. They found that creativity and epistemic curiosity were positively associated with eye movement patterns measured as the number of unique scene regions visited and with the Shannon entropy, which is a way to quantify the complexity of participants' fixation patterns by also considering the frequency with which each region was fixated (Gross et al., 2019). Further, Baranes et al. (2015) reported that higher epistemic curiosity was associated with faster anticipatory saccades towards the location of the expected answers to trivia questions and that trait curiosity was associated with individual variation in curiosity-triggered anticipatory gaze.

The studies reviewed above have used tasks including unrelated stimulus events, such as trivia questions (Baranes et al., 2015) or images unrelated to each other (Gross et al., 2019; Risko et al., 2012). Learning the answers to trivia questions or viewing images unrelated to each other may not be sufficient to trigger further interest in exploring the topics, whereas in natural settings, learners often maintain sustained focus on the same topic over longer periods of time. The ability for sustained investigations is likely to underlie the most important role of curiosity in real life, where acquisition of knowledge and satisfying initial curiosity is not an end point. Instead, the acquired new knowledge could create a positive feedback loop and spark curiosity to learn more about the topic and thus allow people to develop individual interests and skills (Hidi & Renninger, 2019; Murayama, 2022; Ten, Kaushik, Oudeyer, & Gottlieb, 2021). More research is thus needed to understand how epistemic curiosity is related to eye movement patterns and information seeking during natural tasks, specifically in situations requiring sustained focus such as reading.

1.2. N-armed bandit paradigms

To create ecologically valid and dynamic decision-making situations, studies have used *n*-armed bandit paradigms (Speekenbrink, 2022; Steyvers, Lee, & Wagenmakers, 2009; Stojic, Schulz, Analytis, & Speekenbrink, 2020; Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018). These paradigms are well-suited to controlled lab studies but are, at the same time, representative of a broad class of real-world problems. The term 'bandit' refers to a metaphor for a multi-armed slot machine in a casino. The player can choose between multiple options (arms) while trying to maximize reward in a situation where each arm has an independent and unknown payoff distribution. A solution to the bandit problems requires sequential sampling of the options to learn about which arms are better to play (exploration), while also playing the known high-value arms to maximize reward (exploitation). The goal is to improve future decisions by balancing exploration and exploitation. Exploration typically gives a lower immediate payoff but may provide information that improves future choices, leading to better overall performance. The exploitation-exploration dilemma represents a typical conundrum of deciding between something you know and something you do not know, for example, choosing to go to your favorite restaurant instead of trying a new restaurant. Sticking with the old favorite ensures a good meal, but by being willing to explore, you might discover something better.

Optimal solutions to the *n*-armed bandit problems have been extensively studied in machine learning literature (e.g., Gittins, 1979; Kaelbling, Littman, & Moore, 1996). Successful performance in these tasks requires the right balance between exploitation and exploration (Cohen, McClure, & Yu, 2007; Mehlhorn et al., 2015). Studies have shown that the amount of exploration depends on the number of trials available to search the environment (Averbeck, 2015; Wilson, Geana, White, Ludvig, & Cohen, 2014). That is, participants are more likely choose an exploration strategy when they have more trials in which to explore. Importantly, scholars (Kidd & Hayden, 2015; Kobayashi & Kable, 2024) have highlighted the power and flexibility of the *n*-armed bandit tasks to study active information-seeking (i.e., the core behavioral expression of curiosity) because these tasks allow participants to proactively sample and determine which options they wish to explore rather than be concerned with reactive processes towards preselected stimuli.

1.3. The present study

We used an *n*-armed bandit paradigm to investigate the effect of curiosity i) on information seeking, ii) the reading process reflected in eye movements, and iii) the subsequent memory for information. We conceptualized curiosity as motivation toward knowledge and information engagement (epistemic curiosity), distinguishing between ***a dynamic state*** that may fluctuate during information processing (Gruber & Ranganath, 2019; Loewenstein, 1994; Markey & Loewenstein, 2014) and a more stable *individual trait* (Eschmann et al., 2023; Kashdan et al., 2020). Based on this division, we further investigated iv) the relationship of individuals' trait curiosity with both information-seeking behavior and state curiosity evoked by information.

Prior studies have used different variants of *n*-armed bandit tasks to model exploratory and exploitative behaviors (Cohen et al., 2007; Wilson et al., 2014) and their underlying neural mechanisms in information seeking (reviewed in Kobayashi & Kable, 2024). A recent study by Jach et al. (2024) used variants of this task (e.g., tasks called *Investigation* and *Horizon*) to study associations between information seeking and personality traits related to curiosity. To the best of our knowledge, no earlier studies have used the *n*-armed bandit paradigm to investigate how state curiosity affects information seeking, allocation of attention as measured with eye movements during reading, and memory for topics used in the information seeking/reading task.

Compared to trivia question paradigms that allow one-shot information seeking related to isolated events (Fig. 1a), the *n*-armed bandit paradigm increases ecological validity by allowing information seeking and accumulation of information over time (see Kobayashi & Kable, 2024). We used a 3-armed bandit paradigm, to allow curiosity to affect information seeking in an ecologically valid way. Instead of presenting fixed sets of stimuli, our task allowed participants to choose how long they wanted to engage in information seeking. In our task, participants were presented with health-related questions, followed by three boxes presented on a computer screen (Fig. 1b). The boxes contained statements from three different evidence quality categories: scientific and relevant, scientific but irrelevant, or non-scientific but relevant statements related to the question. Participants were asked to select a box, read a randomly selected statement from it, and decide whether they wanted to continue acquiring more information. They could read as many statements as they wished or until the statement selection was exhausted before moving on to answer the question. The task allowed us to quantify switches between the boxes as *exploration* and choosing a statement from the same box as *exploitation* (Mehlhorn et al., 2015).

Besides rating their state curiosity for each statement, we asked participants to provide independent ratings on how useful the statement was for answering the question. The usefulness ratings were collected as a manipulation check to make sure that participants could detect the differences in evidence quality across the three categories. For our task manipulation to work, we expected that evidence quality would be reflected as higher usefulness ratings for the scientific relevant statements. We further asked how confident participants were of their usefulness rating, and how easy the statement was to understand (i.e., clarity). No predictions were made about how the evidence quality would affect confidence and clarity ratings.

Moreover, our analyses controlled for the time-on-task (TOT) and the familiarity with the health-related topics. TOT is known to affect task performance (Hopstaken, Van Der Linden, Bakker, Kompier, & Leung, 2016). As the ratings (particularly curiosity) could vary as a function of TOT, we included it into the analyses to control for the effect of variation in ratings over time. Prior research further indicates that familiar items evoke higher curiosity compared to less familiar items (Berlyne, 1954a) and that metacognition in terms of familiarity is related to curiosity (Kang et al., 2009; Kidd & Hayden, 2015; Metcalfe et al., 2020). To control for the effect of familiarity on curiosity, we included participants' familiarity with the topics of the questions into the analyses.

In addition to the 3-armed bandit task conducted in the lab, our study included an online survey approximately two weeks after the lab task. The online survey was conducted to investigate the relationship between curiosity and memory. To assess participants' self-initiated information seeking behavior outside the lab, we further asked whether they had thought about or searched for more information about the health-related topics after the lab task. To test the association between state curiosity and trait curiosity, the online survey included the Five-Dimensional Curiosity Scale Revised (5DCR) (Kashdan et al., 2020).

1.4. Research questions and hypotheses

Using data from the lab and the online survey, the present study addressed the following research questions and hypotheses.

1.4.1. Research question 1: Does curiosity affect information seeking?

Prior research has proposed that curiosity (as an intrinsic motivation) affects active information sampling by assigning value to specific types of information (Gottlieb & Oudeyer, 2018). In our study, the value of information with respect to answering the questions was related to the evidence quality. In the 3-armed bandit task, participants were supposed to learn about the value (i.e., quality) of the statements by exploring information from three sources. We predicted that state curiosity would be related to evidence quality, such that state curiosity would be rated higher after reading statements from the higher quality categories (*Hypothesis 1*).

According to Vogl et al. (2020), epistemic emotions (e.g., curiosity and surprise) are related to higher exploration of information. We thus predicted that state curiosity would promote information seeking, leading participants to read and exploit more statements from the scientific relevant category during periods of high state curiosity (*Hypothesis 2*). We further expected that curiosity and enhanced information seeking would promote accuracy of answering the questions (*Hypothesis 3*).

1.4.2. Research question 2: Does curiosity affect the reading process?

To investigate the relationship between curiosity and allocation of attention during reading, we tracked participants' eye movements while they were reading the statements. To our knowledge, no prior studies have investigated the effect of curiosity on eye movements during reading. We therefore based our hypothesis on literature that links curiosity and interest to states of attention (e.g., Metcalfe et al., 2020; Unsworth & McMillan, 2013). For example, Unsworth and McMillan (2013) showed that lower topic interest and motivation while reading were related to frequent mind wandering episodes (Smallwood & Schooler, 2015). Mind wandering is related to irregular eye movement patterns that show decoupling from ongoing text processing (Faber, Krasich, Bixler, Brockmole, & D'Mello, 2020; Reichle, Reineberg, & Schooler, 2010), whereas during normal reading, fixation durations vary as a function of word length and frequency, as well as the difficulty of the text (Rayner, 1998, 2009).

We tested the hypothesis that both evidence quality and curiosity would influence eye movements during statement reading. Prior research proposes that curiosity can bias attention and eye movements towards task-related information (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018). Moreover, if participants are curious, they will pay more attention to the statements and have fewer occurrences of mind wandering (Metcalfe et al., 2020). Following these prior studies, we expected that higher state curiosity and reading scientifically relevant information would be associated with a more careful reading of the statements as reflected in participants' eye movement patterns (*Hypothesis 4*). According to prior research (Rayner, 1998, 2009), careful reading is characterized by longer fixation durations, fewer skipping of words and fewer regressions (saccades that move backwards in the text). We predicted that careful reading in our study would be reflected as longer gaze and total fixation duration, lower word skipping rate and longer go-past time (the time spent reading the word and any preceding parts of the text after entering the word but before moving forwards in the text).

1.4.3. Research question 3: Do curiosity and information seeking affect memory for information measured as delayed response consistency?

Stimuli evoking curiosity are thought to receive more attention and are therefore more likely to be remembered (Berlyne, 1954a; Markey & Loewenstein, 2014). In line with this claim, prior research has shown that memory is enhanced for stimuli eliciting higher curiosity (Gruber et al., 2014; Kang et al., 2009; Marvin & Shohamy, 2016). To investigate the relationship between curiosity and information seeking with memory, in the online survey, we showed participants the multiple-choice questions from the lab task and asked them to select the best answer option for each topic. We predicted that delayed response consistency (i.e., choosing consistent responses between the lab task and the survey) as a potential indicator of memory for information would be associated with higher state curiosity (*Hypothesis 5*) and enhanced information seeking (*Hypothesis 6*), reflected as reading more statements, during the lab task.

1.4.4. Research question 4: Is trait curiosity related to state curiosity and information seeking?

Prior research has shown that epistemic trait curiosity correlates with state curiosity (Litman et al., 2005). To obtain a comprehensive view on curiosity, we examined whether state curiosity ratings, collected after reading each statement, correlated with individual trait curiosity (Kashdan et al., 2020). Because our task was related to information seeking and knowledge accumulation, we predicted that epistemic trait curiosity dimensions (joyous exploration and deprivation sensitivity) would correlate with state curiosity (*Hypothesis 7*). Moreover, trait curiosity and particularly deprivation sensitivity, has been shown to predict information seeking in real life (Eschmann et al., 2023). We thus predicted that epistemic trait curiosity would be associated with information seeking during the lab task, reflected as reading more statements and exploring the sources (i.e., switching between the boxes) more frequently (*Hypothesis 8*). Similarly, we expected that epistemic trait curiosity would be associated with increased information processing after the lab task (*Hypothesis 9*). To test this, we asked in the online survey whether the participants had thought about or searched for more information about the health-related topics.

2. Methods

2.1. Participants

Participants (N=60) were recruited through student email lists, noticeboard advertisements, and word of mouth. All participants were Finnish-speaking adults and had normal or corrected-to-normal vision. They did not report any psychiatric or neurological history that could have affected task performance, or any current medication affecting the central nervous system. Each participant gave an informed, written consent via electronic documentation prior to taking part in any research activities and was compensated with a movie ticket. The study was performed in accordance with the Declaration of Helsinki and was approved by the University of Helsinki Ethical Review Board in Humanities and Social and Behavioural Sciences.

One participant was excluded due to not understanding the task instructions. Data from 59 participants (age: 28.9 ± 7.1 years, mean ± SD; 43 female, 14 male, 2 non-binary) were included in the behavioral analysis for the ratings and information seeking behavior. Education was coded into three categories: secondary level education (n=18), lower university level education (n=20), and higher university level education (n=21). Data from 56 participants (Supplementary material, Supplementary Text S1) were included into the eye movement analysis. Three participants had to be excluded from these analyses because of technical issues during recording of their eye-tracking data.

One week after the lab experiment, all participants received an invitation to participate in a follow-up online survey. They were asked to fill in the survey within a week. Those who did not fill in the survey in the requested time were reminded one week after the initial invitation (i. e., two weeks after the lab study), and they also had one week to fill in the survey. The mean delay between the lab and online study was 14 days ± 7 (SD), which is in line with Kang et al. (2009). Fifty-two (age: 29.4 ± 7.4 years, mean ± SD; 39 female, 11 male, 2 non-binary) individuals out of the original sample responded to this survey. They were compensated with an additional movie ticket. Education in this sample was distributed as follows: secondary level education (n=14), lower university level education (n=17), and higher university level education (n=21).

2.2. Procedure

In the 3-armed bandit task, participants were presented with 18 health-related questions one at a time and were subsequently asked to rate their familiarity with the topic of the questions (Fig. 1b). After the familiarity rating, three boxes (A, B, and C) representing the bandit arms were shown on the screen. Each box offered an unknown reward, i.e., contained five statements that were or were not helpful for answering the question. The reward value of the box selection was related to evidence quality of the statements: scientific relevant (SR), scientific irrelevant (SI), and non-scientific relevant (NR). Each box included statements of only one type of evidence quality, but these were shuffled across the questions. The reward associated with the boxes alternated between the questions requiring participants to learn which box contained the highest quality of statements for each question. To learn about the content of the boxes participants were instructed to choose a box and read a randomly picked statement from that box to accumulate knowledge over the rounds (statements) for each question. They could read as many statements as they wished, up to a maximum of 15 per question, before proceeding to answer a question. For each topic, we had one multiple-choice question which was answered by choosing one among four answer options. To analyze participants' information seeking and decision behavior, we considered drawing a statement from the same box as in the previous round as *exploitation* and switching to another box as *exploration* behavior (see Mehlhorn et al., 2015).

After reading a statement, participants were asked to provide ratings on i) how *useful* the statement was for answering the question, ii) how *confident* they were of their usefulness rating, iii) to what extent did the statement evoke their *curiosity* towards the topic of the question, and iv) how easy the statement was to understand (i.e., *clarity*). All ratings were given with a response pad (Cedrus RB-844, Cedrus Corporation, San Pedro, CA, USA) on a scale from 1 (low) to 4 (high). Responses to the questions were also collected with the same response pad including four options. Prior to the experiment, instructions were given, and participants practiced the task.

The online survey was conducted approximately two weeks after the lab task. We showed participants the multiple-choice questions and asked them to select the best answer option for each topic. In addition, the online survey contained questions about whether participants had thought about the health-related topics or searched for more information about them after the lab experiment. The Five-Dimensional Curiosity Scale Revised (5DCR) (Kashdan et al., 2020) was also administered in the online survey (for the 5DCR items and instructions, see Supplementary material, Supplementary Text S2).

2.3. Stimulus materials

The stimulus materials were sourced from Knuuti (2020), a book that contains claims across health-related themes. Eighteen topics were chosen from this book and formulated into health-related questions (e. g., "Aromatherapy relieves anxiety. True or False?"). Each question was associated with 15 statements (270 statements in total). The statements were divided into three quality categories: scientific relevant (SR), scientific irrelevant (SI), and non-scientific relevant (NR). The evidence quality categories were based on two criteria: *relevance* to the question and *authority* defined as the authority's reliability and the extent to which other authorities in the field share the same opinion (Hoeken, Timmers, & Schellens, 2012). Thus, in addition to being relevant to the question, the criteria for defining a statement as a high-quality SR statement were scientific authority's reliability and relevant expertise on the topic (e.g., "According to systematic reviews, lavender aromatherapy reduces anxiety."). The SI statements referred to scientific authority but were irrelevant to the question (e.g., "Based on studies, there are no significant side effects from aromatherapy."). The NR statements were relevant to the question but did not contain references to scientific authority (e.g., "According to the aromatherapist, aromatherapy lifts the mood and relieves anxiety."). Based on the two criteria, a fourth category, non-scientific irrelevant, would have also been possible, but this category was deemed too obvious and for the interest of brevity was not included into the study. Most of the SR and SI statements were sourced from Knuuti (2020), and the NR statements were created by inventing authorities without scientific credibility. All stimuli were presented in Finnish, and their English translations are available at https://osf.io/2zcu6/.

After a preliminary classification of statements, a pilot testing for the evidence quality categories was conducted using REDCap (Harris et al., 2009) electronic data capture tools hosted at University of Helsinki. For the pilot tests, statements were divided into 4 blocks, and the respondents could complete as many of the blocks as they wished. As a result, 12 participants categorized the statements of the 1st block into the three predefined quality categories (SR, SI, or NR) and gave comments related to the statements. The rest of the blocks were categorized and reviewed by 8 participants. These participants did not take part in the study. On average, the participants agreed with the preliminary classification in 85.1% ± 9.1 (SD) of the SR statements, 68.6% ± 14.0 of the SI, and 95.8% ± 5.2 of the NR statements. Pilot participants' agreement with the original categorization varied between categories as indicated by one-way ANOVA ($F(2,51) = 33.18$, $p < .0001$, $\eta^2 = 0.57$). Post hoc Bonferroni corrected t-tests revealed lower agreement for SI statements compared to SR ($p < .001$) and NR ($p < .001$) statements. Moreover, agreement was lower for SR than NR statements ($p = .008$). Based on pilot participants' ratings and feedback, we modified the statements, especially the SI statements, to improve classification reliability.

The mean length of the statements was 13.3 words ± 5.5 (*SD*) for SR, 11.7 words ± 4.3 (*SD*) for SI and 10.8 words ± 3.0 (*SD*) for NR statements. Statement length varied between categories as indicated by one-way ANOVA ($F(2,267) = 7.61$, $p < .001$, $\eta^2 = 0.05$). Post hoc Bonferroni corrected *t*-tests showed that SR statements were longer than SI ($p = .04$) and NR ($p < .001$) statements, while the length of SI and NR statements did not differ ($p = 0.51$). For each word we extracted the word frequencies from Finnish Internet Parsebank (Luotolahti, Kanerva, Laippala, Pyysalo, & Ginter, 2015). For 4.1% of the words, frequency values were not found in the parsebank. These included, e.g., numerals, some words including a non-letter character (e.g., -, %), or abbreviations.

Questions were presented alongside statements to avoid participants having to memorize the questions. However, eye movement analysis focused solely on the statements. The stimuli were presented on a grey background in black font (Courier) with the font size 35 at 90 cm viewing distance. The letter width was approximately 0.4° of visual angle.

2.4. Data acquisition and preprocessing

The stimuli were presented on a HP LA2405x 24" (51.8 x 32.4 cm) monitor with a resolution of 1920 x 1200. Stimulus timing and presentation was controlled using the Presentation software (Version 22.1, Neurobehavioral Systems, Inc., Berkeley, CA, USA). During the experiment, participants were seated in a dimly lit room, approximately 90 cm from the computer screen. Eye movements were recorded with an SR Research EyeLink 1000 remote eye-tracker (SR Research, Ltd., Kanata, Ontario, Canada) at the sampling rate of 1000 Hz. Eye movement data were pre-processed with SR Research DataViewer software (version 4.3.1) including drift corrections and setting up interest areas such that each word in each statement constructed an area of interest. Fixations and saccades were extracted from the raw eye coordinate data using the standard algorithm provided by EyeLink with the following thresholds: minimum velocity of 30°/sec, minimum acceleration of 8000°/sec^2 or a minimum motion of 0.1°.

During data collection, the questions were separated into four blocks, each consisting of 4 or 5 questions. To guarantee the accuracy and precision of the eye-tracking data, the eye tracker was calibrated, and calibration accuracy was validated using a 9-point grid for every participant before each recording block. In a validation step, the calibration was repeated until the average error for all points was less than 1°. The presentation order of the blocks was balanced across participants. Participants were allowed to take breaks between the blocks.

2.5. Statistical analysis

Prior to the statistical analysis, all ratings (state curiosity, usefulness, confidence, clarity, and familiarity with the topic) were centered and z-scored. 1.7% of the trials were excluded due to incorrect button presses. We ran four linear mixed effects models (LMMs) to compare the ratings between evidence quality categories (Appendix A, Table A1-A4). The model for state curiosity addressed *Research question 1* on whether curiosity affects information seeking. The other three ratings were analyzed to make sure that our task manipulation worked, and that participants could detect the differences between the evidence quality categories. The dependent variables (DVs) in these models were the ratings. The independent variables (IVs) were *evidence quality* coded as a factor with three levels (SR, SI, and NR), *familiarity* with the topic and a *trial order* variable (z-scored and centered). Trial order was included to control for the effect of time-on-task (TOT) (see e.g., Hopstaken et al., 2016), and familiarity was included because prior research suggests that familiarity is linked with curiosity (Berlyne, 1954a; Metcalfe et al., 2020). For consistency, we included familiarity to all models for ratings. The random factors were random intercepts for the statements and random slopes for evidence quality at the level of participants. These variables were included to account for variation between the statements and within participant variation in ratings across the evidence quality categories.

To further investigate *Research question 1* on whether curiosity affects information seeking behavior, we ran four additional models. In the first model, the DV was the number of rounds (statements read by the participants) per question (Table A5). As the number of rounds represents count of events, we analyzed it with a generalized linear mixed model (GLMM) using Poisson distribution. The IVs were *mean state curiosity* and *familiarity* with the topics. The random factors were random intercepts for both questions and participants to account for variation between the questions and the individuals. In the second model (LMM), the DV was the proportion of statements read per evidence quality categories (Table A6). The IVs were *evidence quality* and *mean state curiosity*. The random factor was random intercept for participants as the proportion of statements read per category was calculated at the participant level. In the third model, the DV was exploration behavior quantified as switches between the boxes (Table A7), and as a binary measure (yes/no switch), we analyzed it with a GLMM using binomial distribution. The IVs included *evidence quality* and *mean state curiosity*. The random factor included random slopes for evidence quality at the level of participants. In the fourth model, the DV was accuracy of responding to the multiple-choice questions (Table A8). As a binary measure (correct/incorrect), it was analyzed with a GLMM using a binomial distribution. The IVs were *mean state curiosity*, *familiarity,* and *number of rounds* read per question. Number of rounds was included because on top of curiosity and familiarity with the topic, accuracy as the outcome of information search could be affected by the number of statements read. The random factors were random intercepts for both questions and participants.

To address *Research question 2* about the effect of curiosity on the reading process, we ran four models. The eye-tracking data were analyzed at word-level with LMMs. We used four standard DVs to assess reading behavior: 1) gaze duration, 2) go-past time, 3) skipping rate, and 4) total fixation duration. Gaze duration is the summed duration of all fixations on a word during its first-pass reading. Go-past time is the time spent reading the word and any preceding parts of the text after entering the word but before the word is exited to the right for the first time. Skipping rate is a dichotomous variable reflecting whether the word was skipped during first-pass reading or not. Total fixation duration is the summed duration of all fixations landing on a word. The fixation duration measures were log-transformed to normalize the data. The four models included *state curiosity, evidence quality* and their interaction, as well as *word frequency* and *word length* as IVs (Tables A9-A12). We included word *frequency* and *word length* as these are known to affect eye movement metrics (Rayner, 1998, 2009). Word frequencies were used as frequency per 1 million and log-transformed, and both word frequency and length measures were centered. The random factors included random intercepts for participants, statements, and words. We also included random slopes for evidence quality at the level of participants as random factors. Prior to these analyses, 4.3% of the trials were excluded due to calibration inaccuracies detected based on visual inspection of the eye-tracking data. An additional 5.4% of the words were excluded due to missing word frequency information[1]. We also excluded words if the first-pass gaze durations were considered too short (< 50ms) or too long (> 1000ms), comprising of 0.6% and 1.2% of the data, respectively. In addition, we excluded words that were skipped during the first-pass reading from the analyses of fixation duration measures, comprising of 30.9% of the data.

[1] Note that this percentage is higher than the one reported under "2.3. Stimulus materials" due to exclusion of trials and blocks of poor eye-tracking data quality.

We ran one model to investigate *Research question 3* on the effect of curiosity and information seeking on delayed response consistency as a potential indicator of memory (Table A13). In the online survey conducted approximately two weeks from the lab experiment, we showed participants the multiple-choice questions from the lab experiment and asked them to choose the best option out of four possible alternatives. Responses were coded as a binary variable based on whether the responses were consistent or inconsistent between the lab experiment and the survey. Consistent responses were treated as an indicator of participants' memory for the topics, and these responses were used as a DV in the GLMM using binomial distribution. The IVs were *mean state curiosity, familiarity*, *number of rounds,* and the *delay* between the lab experiment and the survey. Familiarity with the topics was added as a control variable and the delay in days (centered) was included because it was likely to affect the memory results. The random factors included intercepts for the questions and participants.

To investigate *Research question 4* on whether trait curiosity is associated with state curiosity and information seeking, participants filled in the 5DCR questionnaire (Kashdan et al., 2020) (Supplementary material, Supplementary Text S2). Prior to the statistical analyses, we reversed the scores for the stress tolerance items, computed the average item scores for the trait curiosity dimensions for each participant, and centered these. The first model was an LMM addressing the relationship between trait and state curiosity. The DV was state curiosity averaged over the statements for each question and for each participant. The IVs were the scores for trait curiosity dimensions (Table A14). The random factor included random intercepts for participants. The next models investigated the associations between trait curiosity and information-seeking behavior. In the second model, the DV was the number of rounds per topic, and it was modeled using a GLMM with Poisson distribution. In the third model, the DV was exploration between the sources, modeled using a GLMM with binomial distribution (Table A15 and A16). The IVs in both models were trait curiosity dimensions, and the random factor included random intercepts per participants. In addition, we investigated the effect of curiosity on information-seeking behavior after the lab experiment. In the online survey, we asked participants whether they had thought about the health-related topics and whether they had searched for information about the topics. The responses were treated as binary variables (yes/no) and were used as DVs in GLMMs with binomial distribution (Table A17 and A18). The IVs were *mean state curiosity*, *familiarity, number of rounds,* and *trait curiosity dimensions* since along with trait curiosity, state curiosity, and familiarity, information seeking behavior in the lab task may also affect information seeking after the lab experiment. The random factor was random intercepts for participants.

The analyses were run using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) for R statistical software (R Core Team, 2025). Separate models were fitted for each DV with the restricted maximum likelihood (REML) criterion. The 95% confidence intervals (CIs) were computed using Wald estimation. The models were created for hypothesis testing and included IVs that we expected would affect the DVs. We used a Likelihood Ratio Test (LRT) using the anova() -function from the stats package (R Core Team, 2025) to test if adding factors improved the model fit. For visualizations, we used the R packages ggplot2 (Wickham, 2016) and sjPlot (Lüdecke, 2025). For GLMMs with counts as DVs (using Poisson distribution), the dispersion ratio was checked using the check_overdispersion() -function from the performance package (Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021). Supplementary material, the analysis codes and data underlying statistics and figures can be found at https://osf.io/pcq3u/.

3. Results

3.1. Ratings and choices

The distribution of ratings across the evidence quality categories are shown in Supplementary material, Figure S1 and the model predicted ratings are shown in Fig. 2 (see also Supplementary material, Table S1 for descriptive statistics and correlations between ratings). Participants (n=59) rated their *state curiosity* (Fig. 2a) the highest after reading SR statements as compared with SI and NR statements (Table A1). State curiosity was also rated higher after reading SI than NR statements (b = -0.27, SE = 0.08, t = -3.49, $p < .001$) when SI was used as the baseline in the analysis. State curiosity ratings were sensitive to the time-on-task (TOT, Table A1, Supplementary material, Figure S2a) with decreasing curiosity over time. Familiarity with the topics was also associated with higher state curiosity. Curiosity ratings also correlated positively with usefulness and familiarity ratings (Supplementary material, Table S1).

The *usefulness* (Fig. 2b) of the statements for answering the questions was rated the highest for SR statements (Table A2), indicating that our task manipulation worked as the highest quality statements were rated as the most useful. The usefulness ratings did not differ between SI and NR statements, nor did the usefulness ratings vary over time or familiarity with the topics (Supplementary material, Figure S2b).

Confidence (Fig. 2c) of the usefulness ratings were the lowest for SI statements (Table A3). Confidence ratings did not differ between SR and NR statements and did not vary over time (Supplementary material, Figure S2c). However, confidence ratings were higher for familiar topics and were also correlated with clarity ratings (Supplementary material, Table S1).

Clarity (Fig. 2d) of the statements was rated the highest for NR statements (Table A4). Clarity did not differ between SR and SI statements. However, the clarity ratings increased with TOT (Supplementary material, Figure S2d) possibly indicating learning over the course of the experiment. Moreover, statements related to familiar topics were rated higher on clarity. Clarity further correlated with familiarity ratings (Supplementary material, Table S1).

3.2. Information seeking

Participants made on average, 4.9 ± 3.2 (SD) rounds per question, with a substantial individual variability (Supplementary material, Figure S3). The majority of the participants read at least 50 statements but only a few read more than 150 statements, while none read all the 270 statements (Supplementary material, Figure S4). Both state curiosity (p = .077) and familiarity with the topic (p = .061) affected the number of rounds per question at a trend level but did not reach significance (Table A5). As the number of rounds represents count of events, it was analyzed with a GLMM using Poisson distribution. The dispersion ratio of the model was 0.59, indicating no overdispersion.

The proportion of statements read varied between the evidence quality categories (Fig. 3a). Participants read more SR statements than SI or NR statements (Table A6), with no difference between the proportion of SI and NR statements read. Moreover, state curiosity was associated with the proportion of statements read at a trend level (p = .060) but did not reach significance.

To analyze exploration in terms of how often participants switched between the boxes (see Fig. 1b), the first statements per each question were excluded from the data as switching between the sources was not possible during the first trial. Overall, participants explored the sources 70% ± 46 (SD) of the time, indicating that they were in an exploration mode most of the time. Although not included to the initial model, we tested whether time-on-task (TOT) affected switching between the boxes, because prior research indicates that exploration depends on the number of trials available to explore the environment (Averbeck, 2015; Wilson et al., 2014). The model fit improved when TOT was included in the model ($X^2(1) = 74.76$, $p < .001$). Exploration was lower for relevant scientific than non-scientific statements as indicated by a lower probability of switching to the SR than NR statements from other evidence quality categories (Fig. 3b, Table A7). The participants thus exploited SR statements more than NR statements. There was no difference in switching to the SI and NR statements from other evidence quality

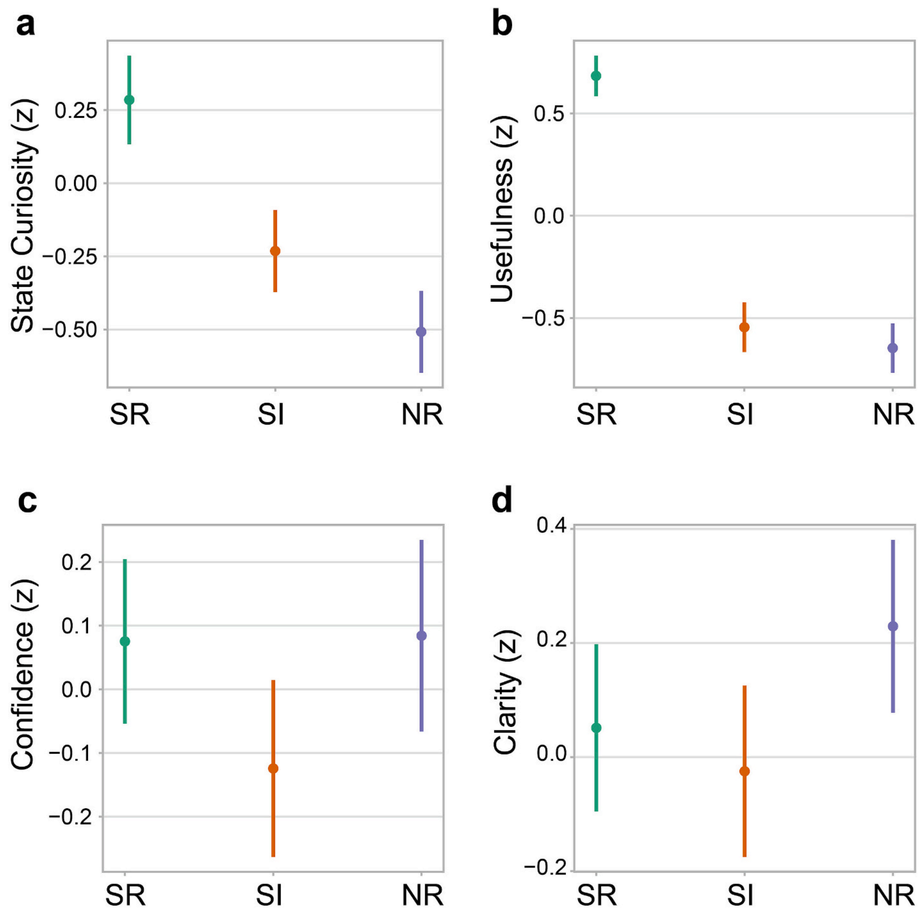


Fig. 2. *Rating results* Model predictions for a) state curiosity, b) usefulness, c) confidence, and d) clarity ratings across the evidence quality categories: SR = scientific relevant (green), SI = scientific irrelevant (orange), NR = non-scientific relevant (purple), controlling for the effect of time on task and familiarity with the topics. Error bars denote 95% confidence intervals (CI), n=59.

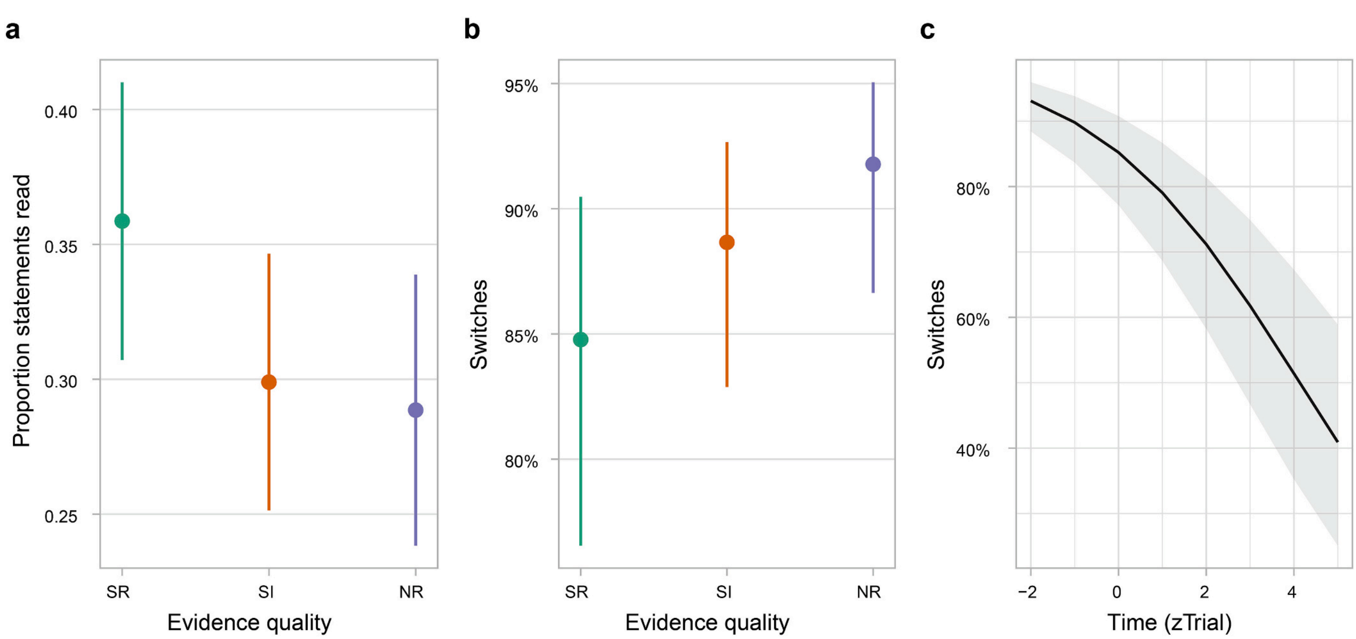


Fig. 3. *Information seeking results* a) Predicted proportion of statements read varied across the evidence quality categories and was highest for the SR statements. Predicted probability of exploration (i.e., switches between the boxes, see Figure 1b) indicated that exploration varied across b) the evidence quality categories and was lower for SR than NR statements and c) decreased over time-on-task (p < .001). SR = scientific relevant (green), SI = scientific irrelevant (orange), NR = non-scientific relevant (purple). Error bars and shading denote 95% confidence intervals (CI), n=59.

categories. State curiosity was not associated with exploration. However, exploration decreased over time (Fig. 3c, Table A7).

Participants answered 74% ± 44 (*SD*) of the multiple-choice questions correctly. Correct answers to the questions were associated with a higher number of rounds per question (Table A8) but state curiosity and familiarity did not affect the accuracy of answering the questions.

3.3. Eye movements during reading of the statements

The non-transformed descriptive statistics for the eye movement variables are presented in Table 1. To obtain singularity for the gaze duration analysis, we excluded the random slopes for evidence quality at the level of participants from the random part of the model. Evidence quality and state curiosity had no effect on gaze duration (Table A9). However, longer gaze duration was associated with longer words and shorter gaze duration was associated with higher frequency words. Evidence quality and state curiosity had no significant effect on go-past time (Table A10) but longer go-past times were associated with longer words. Further, these analyses showed several trends that did not reach significance, suggesting longer gaze durations for SR than NR statements (p =. 076) and longer go-past times for SR than SI statements ($p = .053$). Further, evidence quality interacted with curiosity at a trend level for both gaze duration ($p = .067$) and go-past time ($p = .074$), suggesting stronger effects of curiosity for SI than SR statements.

For skipping rate, main effects for evidence quality indicated higher skipping rate during the first-pass reading of SR statements compared to SI and NR statements (Fig. 4a, Table A11). The skipping rates did not differ between SI and NR statements when SI was used as the baseline [b = -0.03, SE = 0.07, z = -0.45, p = .655]. Moreover, the analysis indicated that higher state curiosity was associated with lower skipping rate (Fig. 4b, Table A11). Finally, skipping rate was associated with both word length and word frequency: High frequency words were skipped more often than low frequency words, and shorter words were skipped more often than longer words. Evidence quality further interacted with curiosity at a non-significant trend level ($p = .054$), suggesting that with higher state curiosity skipping rate decreased for SR but not for NR statements.

The analysis for total fixation duration showed an interaction between state curiosity and evidence quality (Table A12). Although higher state curiosity was associated with longer total fixation duration irrespective of evidence quality, curiosity had stronger effects on total fixation duration for SI and NR statements as compared with SR statements (Fig. 5, Table A12). Total fixation duration was longer for SR compared to NR statements, and NR statements received significantly shorter total fixation duration than SI statements [b = -0.07, SE = 0.03, t = -2.19, p = .028] when SI was used as a baseline. Separate analyses for the evidence quality categories showed that increases in total fixation duration with state curiosity were steeper for both SI [b = 0.05, SE = 0.007, t = 6.21, p < .001] and NR [b = 0.06, SE = 0.01, t = 7.72, p < .001] than for SR [b = 0.02, SE = 0.01, t = 3.43, p = .001] statements. In other words, when state curiosity was low, evidence quality had larger impact on total fixation duration, whereas the differences in total fixation duration between the evidence quality categories were smaller when state curiosity ratings were high. In addition, we observed effects of word frequency and word length for total fixation duration (Table A12), with longer total fixation duration for longer words and shorter total fixation duration for higher frequency words.

Table 1
Non-transformed means and standard deviations (in parenthesis) for the eye movement variables across evidence quality categories. n=56

	SR	SI	NR
Gaze duration (ms)	280 (152)	268 (141)	269 (147)
Go-past time (ms)	801 (1499)	735 (1452)	647 (1218)
First-pass skipping rate	0.32 (0.47)	0.32 (0.47)	0.29 (0.45)
Total fixation duration (ms)	604 (497)	569 (481)	529 (458)

SR = Scientific Relevant; SI = Scientific Irrelevant; NR = Non-Scientific Relevant; ms = milliseconds

3.4. Memory for topics after the lab experiment measured as delayed response consistency

In the online survey, we presented participants with the multiple-choice questions from the lab experiment and asked them to choose the best answer option. Choosing consistent responses with the lab experiment was treated as an indicator of participants' memory for the topics. 74% ± 44 (*SD*) of the responses in the online survey were consistent (while 73% ± 45 (*SD*) of the responses were correct). We found no effect of state curiosity on the delayed response consistency. However, choosing consistent responses was positively associated with the number of rounds during the lab task (Table A13; Supplementary material, Figure S5), while controlling for the delay (in days) from the lab experiment.

3.5. The effect of trait curiosity on state curiosity and information seeking

To investigate the relationship between state and trait curiosity, we pooled data from state curiosity ratings across the statements for each question and compared these with individual mean scores for trait curiosity dimensions from the 5DCR (Kashdan et al., 2020). The results showed that thrill seeking correlated negatively with state curiosity (Table A14, Supplementary material, Figure S6a), while other trait curiosity dimensions were not associated with state curiosity.

In addition, we tested whether trait curiosity was associated with information seeking in the lab task in terms of the number of rounds and exploring the sources, i.e., switching between the boxes (calculated as individual sum of switches per each question). Trait curiosity did not explain information seeking in the lab task, as the trait curiosity dimensions were associated with neither the number of rounds (Table A15) nor with exploring the sources (Table A16). Only overt social curiosity was positively related to number of rounds at a non-significant trend level ($p = .069$). As both variables represent counts of events, we analyzed them with GLMMs using Poisson distribution. The dispersion ratios were 0.64 and 0.78, respectively for the number of rounds and for exploration of the sources. Both values were < 1, indicating no overdispersion.

To investigate whether curiosity was related to information seeking after the lab task, we asked in the online survey, whether participants had thought about any of the health-related topics covered in the lab experiment and whether they had searched for information about these topics after the experiment. 48% ± 50 (*SD*) of the participants reported thinking about the topics and 15% ± 36 (*SD*) of them had searched for information about some of the topics. We analyzed whether thinking and searching for information were related to participants' state and trait curiosity levels, familiarity with the topic, and the number of rounds per topic in the reading task. Neither state curiosity, familiarity, nor the number of rounds were related to thinking about the topics (Table A17). However, both joyous exploration and deprivation sensitivity of trait curiosity dimensions were positively associated, while thrill seeking was negatively associated with thinking about the topics (Table A17, Supplementary material, Figure S6b-d). Searching for information about the topics after the lab experiment was not associated with curiosity, familiarity, nor the number of rounds per topic (Table A18).

4. Discussion

Experimental studies of curiosity (Baranes et al., 2015; Gruber et al., 2014; Jach, DeYoung, & Smillie, 2022; Kang et al., 2009; Marvin & Shohamy, 2016) have typically used tasks in which participants are asked to rate their curiosity about unrelated stimulus events, such as

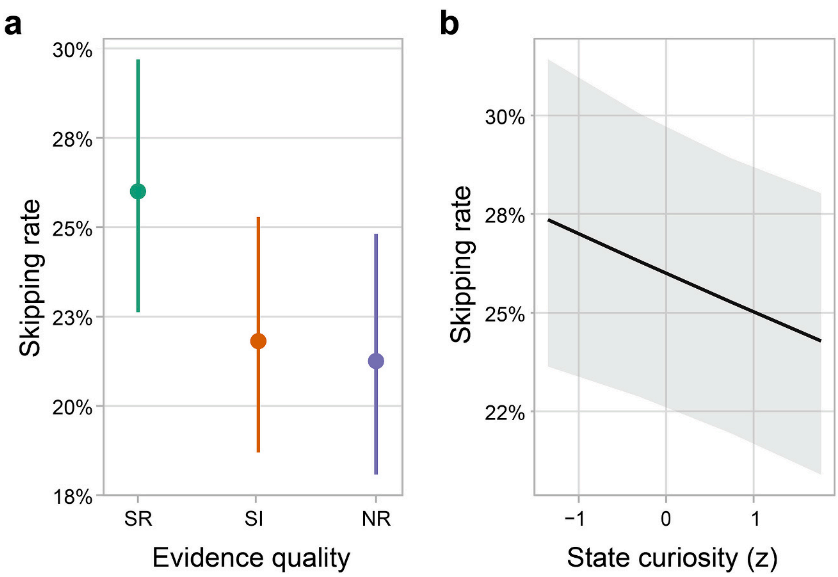


Fig. 4. *Skipping rates* Predicted skipping rates a) varied across the evidence quality categories and were highest for the SR statements and b) decreased with increasing state curiosity ($p = .018$). SR = scientific relevant (green), SI = scientific irrelevant (orange), NR = non-scientific relevant (purple). Error bars and shading denote 95% confidence intervals (CI), n=56.

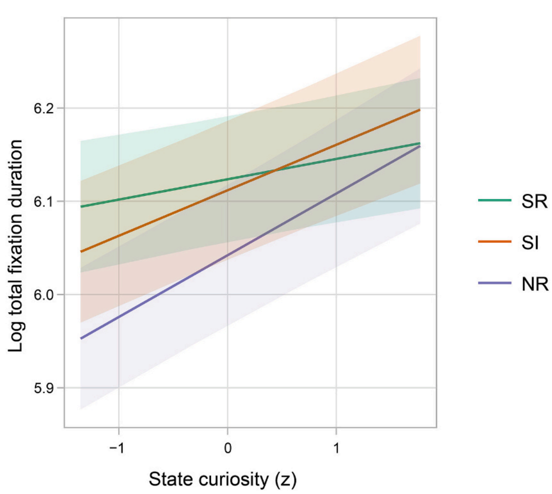


Fig. 5. Total fixation duration Total fixation duration as a function of state curiosity across the evidence quality categories. SR = scientific relevant (green), SI = scientific irrelevant (orange), NR = non-scientific relevant (purple). Shaded areas denote 95% CI, n=56.

trivia questions (Figure 1a). The present study used an adaptation of the n-armed bandit paradigm (Fig. 1b) (Speekenbrink, 2022; Steyvers et al., 2009; Stojic et al., 2020; Wu et al., 2018), in which the “arms” (sources) varied in the quality of evidence across three categories: scientific relevant (SR), scientific irrelevant (SI), and non-scientific relevant (NR). Participants were free to choose how many statements to read per question, which allowed curiosity to affect information seeking over time and in a more ecologically valid way compared to presenting the participants with fixed sets of stimuli per condition. Previous studies have used different variants of n-armed bandit tasks to investigate information-seeking behaviors (Cohen et al., 2007; Wilson et al., 2014) and their neural correlates (see for review Kobayashi & Kable, 2024). Jach et al. (2024) further showed that shared variance across tasks related to information demand (including variants of the bandit tasks) was related to personality traits reflecting curiosity. However, no earlier study has used the n-armed bandit task to investigate how evidence quality and state curiosity affect information seeking.

Moreover, prior studies investigating the effect of curiosity on attention and eye movements are scarce (Baranes et al., 2015; Gross et al., 2019; Risko et al., 2012) and have used either trivia questions or images unrelated to each other. The effects of curiosity on eye movements during natural tasks, requiring sustained focus on the same topic (e.g., reading), are therefore not well understood. Using the n-armed bandit paradigm and an online survey after the lab experiment, our study provides novel evidence about *the effects of evidence quality and state curiosity i) on information seeking, ii) the reading process reflected in eye movements, iii) the subsequent memory for information.* The study further assessed *iv) the relationship of individuals’ trait curiosity with information-seeking behavior and state curiosity evoked by information.*

4.1. *Evidence quality, familiarity with the topics, and time-on-task affected ratings*

Regarding *Research question 1* on whether curiosity affects information seeking, we found that participants rated their state curiosity highest after reading SR statements. This is in accordance with *Hypothesis 1* and prior research (Gottlieb & Oudeyer, 2018), indicating that curiosity affects information seeking by assigning value to information. In our experiment, SR statements had the highest value for answering the questions. SR statements were also rated the highest on usefulness, confirming that the evidence quality manipulation was successful as participants were able to detect the high-quality from the lower-quality evidence.

Results for the other ratings showed that confidence about the usefulness of the statements was the lowest for SI statements that were also rated the lowest on clarity, suggesting that SI statements were the most difficult to comprehend. NR statements were rated highest on clarity and were thus the easiest to comprehend. The clarity ratings did not differ between SI and SR statements, suggesting that scientific content was more difficult to comprehend compared to the non-scientific statements.

Prior research proposes that curiosity is based on metacognition, such as detection of knowledge gaps and a feeling of knowing (Loewenstein, 1994; Metcalfe et al., 2020). To control for participants’ prior knowledge with the health-related topics, we included familiarity with the topics to the analyses for ratings. In line with previous research (Berlyne, 1954a; Loewenstein, 1994; Metcalfe et al., 2020), we found that curiosity was rated higher for familiar topics. We further found that familiarity with topics was associated with higher ratings for both confidence and clarity. Prior research (Kang et al., 2009) indicates non-linear relationships between curiosity and confidence, whereby curiosity is highest at intermediate levels of confidence. Similar results have also been observed among 7- to 8-month-old infants whose attention peaked for moderately surprising stimuli in relation to previously presented stimuli, i.e., for stimuli falling in a "Goldilocks" range (Kidd et al., 2012; Kidd et al., 2014). However, our analyses did not test non-linear effects, and curiosity was not correlated with confidence.

In trivia paradigms, the answers to trivia questions provide information that is not connected to an external reward. Based on these results, definitions of curiosity have emphasized the viewpoint that curiosity is non-instrumental and reflects intrinsic motivation without any apparent external reward (e.g., FitzGibbon, Lau, & Murayama, 2020; Gottlieb & Oudeyer, 2018; Loewenstein, 1994). In contrast, we found that state curiosity correlated with usefulness ratings, reflecting higher external gain and that information in our task had instrumental value for guiding future choices. Participants were thus more likely to feel curious about a statement if they perceived it to be useful for answering the question (see also Dubey, Griffiths, & Lombrozo, 2022). Our results showed that both usefulness and state curiosity were strongly correlated and both were rated highest for SR statements. The co-occurrence of high state curiosity and usefulness ratings thus reflects a property of our task, where information was acquired to answer questions, that differed from the trivia paradigms, where participants request information for its own sake.

We also found important differences between the curiosity and usefulness ratings. As shown before (Berlyne, 1954a), we found that state curiosity increased with familiarity, whereas the usefulness ratings were not related to familiarity. Moreover, the usefulness ratings did not vary over time-on-task (TOT), but state curiosity declined over time. Previous research has shown that with increasing TOT, behavioral performance and pupil diameter declines (van den Brink, Murphy, & Nieuwenhuis, 2016) and engagement with the task decreases while experienced fatigue increases (Hopstaken et al., 2016). These studies indicate a decline in attentional performance over time. The decrease in state curiosity ratings over time observed here is in accordance with previous findings and further supports the link between curiosity and attention, indicating that curiosity was sensitive to increased mental fatigue. Moreover, we found that clarity ratings increased over time, suggesting a possible learning effect over the course of the task.

4.2. Evidence quality affected information seeking

To address the effect of curiosity on information seeking (*Research question 1),* we used four measures to quantify the information-seeking behavior in the 3-armed bandit task. To allow curiosity to affect information seeking in a naturalistic way, participants were free to decide on how many statements to read before answering the questions. First, we analyzed *the number of rounds* (statements) participants read and found that familiarity with the topic and state curiosity were associated with the number of rounds at a trend level, suggesting a higher number of rounds when familiarity and state curiosity were lower. The effect of curiosity was also opposite to what we predicted as participants were reading fewer statements when their state curiosity was high. However, no strong conclusions can be made based on these results as the effects were small and non-significant.

Second, we analyzed the *proportion of statements read per evidence quality category* and found that participants read more of the SR statements than the lower quality statements, indicating that they learned to exploit the high-quality sources more than the other sources. Moreover, state curiosity was positively associated with the proportion of statements read at a trend level. As the proportion of statements read was analyzed at participant level, the non-significant effect of curiosity could be related to limited statistical power.

Third, we quantified switches between the sources (boxes) as *exploration* and choosing a statement from the same box as during previous round as *exploitation*. We found that participants spent most of the time (70%) in an exploration mode and that evidence quality was related to exploration. That is, exploration was lower after reading SR than NR statements. In other words, participants exploited SR statements more than NR statements. Contrary to our expectations, state curiosity was not related to exploration of the sources, but we found that exploration decreased over TOT. This is possibly due to the finite horizon of our bandit task (i.e*.,* the maximum of 15 statements per question) because earlier studies have reported that exploration increases with the number of trials in a bandit task (Averbeck, 2015; Wilson et al., 2014).

Finally, we analyzed the *accuracy* of answering the questions. The results indicated that reading more statements was beneficial for the task performance, as the number of rounds was associated with higher accuracy in answering the questions. Curiosity and familiarity with the topics did not affect accuracy.

Overall, we failed to find evidence for *Hypothesis* 2 as state curiosity did not significantly affect information seeking quantified as the number of rounds, proportion of statements read, and exploration of the sources. However, evidence quality was associated with information seeking as participants read and exploited the highest quality statements the most. Reading more statements further translated into better accuracy while curiosity did not affect accuracy. *Hypothesis 3* was thus only partially supported as information seeking in terms of the number of rounds, but not curiosity, was related to better accuracy. Taken together, evidence quality but not curiosity affected information seeking in our task and only information seeking during the task was beneficial for accuracy.

4.3. Eye movements during reading reflect evidence quality and state curiosity

Prior research implicates eye movements as informative of overt information sampling and choice behaviors (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018). We collected participants' eye movements to assess whether curiosity affected the reading process (*Research question 2*) while reading the statements in the 3-armed bandit task. We hypothesized that both high-quality information and higher curiosity would be associated with a careful reading strategy. Careful reading is characterized by longer fixation durations, fewer regressions back in the text, and fewer skipping of words (Rayner, 1998, 2009), whereas lapses in attention, i.e., mind wandering episodes, are related to irregular eye movement patterns that show decoupling of attention from ongoing text processing (Faber et al., 2020; Reichle et al., 2010). In support of our *Hypothesis 4*, we found that higher state curiosity was associated with lower skipping rate of words during the first-pass statement reading, reflecting more careful reading with increasing state curiosity. Our results further showed that skipping rate was the highest for SR statements during the first-pass reading, but the total fixation duration on words was longer for both SR and SI than NR statements. Thus, the frequent word skipping while initially reading the SR statements was later compensated by more intensive rereading of these statements. Taken together, the results indicated that the scientific statements were read more carefully than the non-scientific statements when both were relevant for the topic.

Importantly, we observed an interaction between state curiosity and evidence quality in total fixation duration on words. This interaction indicated that whether participants were in a curious state had less impact on reading of the SR statements, possibly because they could detect the relevance of these statements. However, increasing state curiosity had stronger effects on total fixation duration for lower evidence quality, SI and NR, categories. Contrary to *Hypothesis 4*, the low-quality statements were both read more carefully with increasing curiosity. This possibly helped participants to engage with low-quality information to determine the credibility and usefulness of that information. Alternatively, it could reflect participants willingness to use their resources on task-irrelevant information, when their state curiosity was high.

Additionally, the eye movement analyses indicated several trends that did not reach significance but supported *Hypothesis 4*. For example, evidence quality interacted with state curiosity in the analyses for gaze duration, go-past time, and skipping rate. The results suggested that with increasing state curiosity the skipping rates decreased for SR but not for NR statements. Moreover, the increase in both gaze duration and go-past time with increasing curiosity was steeper for SI than SR statements.

As a validity check for the eye movement analysis, our results confirmed the standard word frequency and word length effects (Rayner, 1998, 2009). Longer gaze duration was associated with longer words, and shorter gaze duration was associated with higher frequency words. Longer go-past times were associated with longer words. Both word length and word frequency affected skipping rate, whereby high frequency words were skipped more often than low frequency words and shorter words skipped more often than longer words. In addition, longer total fixation duration was associated with longer words and with lower frequency words.

4.4. Delayed response consistency as a potential indicator of memory for the topics was related to information seeking during the task

Curiosity during the initial exposure to stimuli is typically associated with enhanced memory for the items evoking higher curiosity (Gruber et al., 2014; Kang et al., 2009; Marvin & Shohamy, 2016). As we did not provide feedback about the correct answers to the questions in the lab task, unlike other studies (Kang et al., 2009), we could not ask participants to recall the correct answers in the online survey. Instead, to address whether curiosity affected memory for information (*Research question 3*), we asked participants to choose, out of four possible alternatives, the option they thought best responded to the question. The delayed response consistency was treated as an indicator of their memory for the topics. Contrary to *Hypothesis 5*, we did not find any effect of curiosity on the delayed response consistency. For example, Gruber et al. (2014) conducted their memory test directly after scanning participants' brain activity or within 24 h from the behavioral tasks, while the delay in our study was much longer (14 days ± 7 (*SD*)). However, Kang et al. (2009) had their follow-up study within 11–16 days and nevertheless showed that curiosity during the initial session strongly affected recall of the answers to the questions that were initially guessed wrong. The discrepancy between participants' answers and the correct responses could have boosted the memory effects. As we did not provide feedback about the answers in our study, we could not conduct similar analyses to Kang et al. (2009). The delayed response consistency largely reflected accuracy for answering the questions, which was neither affected by curiosity. These methodological aspects may thus explain the lack of association between curiosity and memory here.

Instead, information seeking, measured as the number of rounds (statements read) during the task, was positively associated with the delayed response consistency. *Hypothesis 6* was thus supported, indicating an association between the delayed response consistency and enhanced information seeking during the lab task. These results further highlight that reading more statements during the lab task was beneficial for task performance and for memorizing the topics after the lab experiment. However, one may also argue that giving the same answer on both occasions may not be indicative of memory representation constructed during the experiment. For example, it may be the case that the participants already had a stable existing knowledge representation prior to the study. Alternatively, they may have either reconstructed the same inference or simply guessed in similar way during the second testing round independent of the first round.

4.5. Trait curiosity was associated with state curiosity and thinking about the topics after the lab task

Previous research has shown that differences in distribution parameters of states correlate with individual differences in questionnaire assessments of traits (Fleeson & Gallagher, 2009; Fleeson & Jayawickreme, 2015; Litman et al., 2005; Silvia, 2008). This study investigated whether trait curiosity is related to state curiosity and information seeking (*Research question 4*). To address this question, mean state curiosity ratings, collected after reading each statement, and information seeking behavior were correlated with individual trait curiosity measured with the 5DCR (Kashdan et al., 2020).

We found that state curiosity was negatively associated with thrill seeking, indicating a relatively small impact of trait curiosity on state curiosity ratings. *Hypothesis 7* was thus only partially supported as state curiosity was associated with thrill seeking but not with epistemic trait curiosity dimensions (joyous exploration and deprivation sensitivity) as we had expected. The negligible association between trait and state curiosity could be related to our task where state curiosity measures were specific to the task and statements while the trait curiosity dimensions measured general individual dispositions.

We further tested the relationship between trait curiosity and information seeking in the 3-armed bandit task. Contrary to *Hypothesis 8*, we found that information seeking in the bandit task, measured as the number of rounds and exploration of the sources, was not related to trait curiosity. We also predicted (*Hypothesis 9*) that epistemic trait curiosity would be associated with increased information processing, i.e., thinking and searching for information about the health-related topics after the lab experiment. In accordance with our hypothesis, we found that both joyous exploration and deprivation sensitivity correlated positively with thinking about the topics after the experiment. However, thinking about the topics did not translate to searching for information on them as trait curiosity was not associated with searching for information after the lab experiment, suggesting that individual curiosity alone was not sufficient to motivate people to engage in information seeking in addition to thinking about them after the lab experiment. These analyses further showed that familiarity with topics, information seeking and state curiosity during the lab task were not associated with information processing after the lab experiment, suggesting that familiarity with the topics did not explain why participants did not engage in searching for more information about the topics.

The difference in results for information-seeking behavior in the lab task and after it may also reflect discrepancies between laboratory and real-world contexts. Even though our paradigm aimed to enhance ecological validity by allowing participants to choose how many statements to read per topic, the experiment was still conducted in controlled laboratory settings where participants were possibly concerned about their performance. The setting may have obscured individual differences and explain why we did not find an association between trait curiosity and information seeking in the lab but found that trait curiosity was related to thinking about the topics after the lab task.

We further found that deprivation sensitivity was a stronger predictor of thinking about the topics after the lab session than joyous exploration. According to the information gap theory (Loewenstein, 1994), curiosity can be viewed as a deprivation phenomenon (Litman, 2008) that is characterized by experienced discomfort until information gaps are resolved. Prior research (reviewed in Lydon-Staley et al., 2021) suggests that deprivation curiosity does not motivate people to learn just for fun but includes an element of compulsiveness. This compulsiveness and discomfort may explain why deprivation sensitivity was a stronger predictor in our study. Our results are in line with prior research (Lydon-Staley et al., 2021), showing that deprivation curiosity was associated with an increased tendency to return to previously visited concepts, when participants searched information from Wikipedia articles. Moreover, Eschmann et al. (2023) found that self-reported estimates of COVID-19-related information seeking in real life correlated with deprivation sensitivity but not with joyous exploration.

Intriguingly, we also found that thrill seeking was negatively correlated with thinking about the topics after the lab experiment as well as memory for the topics and state curiosity. Thrill seeking reflects a dispositional tendency to actively seek arousal and includes items such as "Risk-taking is exciting for me.", and "When I have free time, I want to do things that are a little scary." (Kashdan et al., 2020). Thrill seeking has also been linked with disadvantageous outcomes such as unwanted negative emotional experiences and impulsive decision-making (MacPherson, Magidson, Reynolds, Kahler, & Lejuez, 2010). Participants scoring high on this trait possibly found the task and the health-related statements boring and not engaging.

4.6. Limitations and future directions

Epistemic curiosity is related to metacognition about one's current knowledge, such that curiosity peaks with intermediate levels of familiarity and confidence (Kang et al., 2009; Kidd & Hayden, 2015; Metcalfe et al., 2020). We found that familiarity with the topics was related to curiosity and clarity ratings, and confidence about the usefulness ratings. However, we did not ask participants to rate their confidence about the answers to the questions. Thus, we could not test the relationship between curiosity and confidence in a way that would allow comparison of the results with prior studies. Nevertheless, we could show a relationship between metacognition and curiosity with respect to familiarity. Moreover, we did not disclose the answers to the questions to participants, and therefore our results were not directly comparable with the trivia paradigms measuring how curious people were to know the answers. In retrospect, we realize that we could have disclosed the answers for this reason and to investigate the relationship between curiosity and memory in more straightforward fashion instead of measuring delayed response consistency between the answers from the experiment and survey as a potential indicator of memory. However, the focus in our task was on information seeking and we did not want to interfere with this process by guiding participants to select the SR sources. Therefore, we initially chose not to disclose the answers. We thus suggest that future studies adopting this kind of paradigm, should alter the way in which confidence was investigated and consider revealing the answers to the questions perhaps after the experiment and ask how curious people are to know the answers.

To obtain a more comprehensive view on curiosity, we investigated the association between state curiosity and trait curiosity. Compared to previous studies (e.g., Fleeson & Gallagher, 2009), our sample size was limited for a between-person analysis looking at the relationships between individual trait curiosity and state curiosity. Nevertheless, we think it was important to include a measure of trait curiosity to understand to what degree state and trait curiosity correlate in this type of information seeking task. Moreover, state curiosity was asked with one item, while the trait measures included in the 5DCR provide a more fine-grained insight into different motivators underlying curiosity. The trait measures have also been through many stages of validation. As suggested by Jach et al. (2022), future investigations of emotions and information seeking should improve the assessment of state emotions by creating latent measures across several different emotion items. These items could, for example, tap into how state emotions shift over the course of an experiment from pre-information to post-information emotions. Moreover, recent studies (Dawson et al., 2024; Jach et al., 2024) have combined curiosity related measures from personality psychology and cognitive science to create latent measures of curiosity. Future studies could adopt similar approaches to investigate convergence of curiosity constructs across disciplines to improve the validity of curiosity measures and their ability to predict important life outcomes such as learning and academic achievement.

The online survey results showed that trait curiosity dimensions reflecting epistemic curiosity, i.e., joyous exploration and deprivation sensitivity, were related to thinking about the topics after the experiment. However, trait curiosity was not related to searching for information after the experiment, indicating that this effect did not translate into information seeking in real life. An interesting topic for future studies would be to find out the reasons for this discrepancy and what kind of self-motivation, on top of thinking about the topics, is required for participants to engage in information seeking.

Finally, curiosity is a broad term that overlaps with the concept of interest. The use of different concepts is partly related to research fields (Murayama, 2022), i.e., research on interest has emerged from applied research fields such as educational and social psychology while curiosity research has become popular among cognitive science and neuroscience. While there are no clear definitions of these concepts (Hidi & Renninger, 2019; Murayama, 2022), scholars have suggested some distinctions between these two phenomena. For example, curiosity is related to a specific knowledge gap while interest can be more broadly defined as intrinsically motivated engagement with an activity or knowledge (Markey & Loewenstein, 2014; Pekrun, 2019). In line with this distinction, research on curiosity has been mainly interested in short-term information-seeking behavior associated with momentary feelings of curiosity on a trial-by-trial basis, whereas research on interest has been concerned in how and why people can sustain long-term learning (Murayama, 2022). Furthermore, Pekrun (2019) has proposed that curiosity and interest can be treated in conceptually parallel ways while the state-trait distinction is more critical for distinguishing between momentary and more enduring forms of both curiosity and interest. To bridge gaps between different research traditions and concepts, our study combined trial-by-trial level measures, with more sustained information-seeking behavior and trait measures. Instead of using unrelated stimulus events, our paradigm mimicked an ecologically valid and dynamic information-seeking situation and focused on the effect of curiosity in a continuous task that involved learning about a topic over trials, also conceptualized as interest (see Dan, Leshkowitz, Livnat, & Hassin, 2025). In addition to prior theoretical (Murayama, 2022) and computational (Lydon-Staley et al., 2021) work, the current study was an experimental attempt to combine the different aspects of curiosity and interest research.

5. Conclusion

This study investigated the effect of curiosity in reading and information seeking when participants could choose information either by exploring different quality statements or exploiting statements from the same quality category. The setup mimicked real world information seeking in an experimentally controlled environment. We found that state curiosity was higher after reading good quality information, and that higher state curiosity was associated with lower skipping rate of words and higher total fixation duration on words. However, the effect of curiosity on total fixation duration depended on evidence quality. Increasing curiosity had stronger effects on total fixation duration for low-quality statements, suggesting that these statements were read more carefully with increasing curiosity. Moreover, evidence quality was related to information seeking behavior as participants were reading and exploiting the higher quality statements more frequently than the low-quality information. State and trait curiosity did not explain information seeking in the lab task, but behavior in the task was associated with accuracy of answering the questions and memory for the topics after the experiment. We also found that epistemic trait curiosity was related to thinking about the topics after the experiment. Taken together, higher curiosity was related to increased engagement with information.

CRediT authorship contribution statement

Jaana Simola: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Hanna Julku:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Caitlin Dawson:** Writing – review & editing, Methodology, Conceptualization. **Sebastian Österman:** Writing – review & editing, Software, Investigation. **Kaisa**

Elovaara: Writing – review & editing, Investigation. **Tuomo Häikiö:** Writing – review & editing, Formal analysis.

Declaration of competing interest

The authors have nothing to declare.

Acknowledgements

This work was supported by the Strategic Research Council at the Academy of Finland [grant numbers: 335236 and 358272]. The funding source was not involved in the study design, collection, analysis and interpretation of data, writing the report and decision to submit the article for publication. The authors thank Annika Svedholm-Häkkinen (University of Helsinki) and Mika Hietanen (Lund University) for help in defining the evidence quality categories based on argumentation theory.

Table A1
State curiosity: Fixed and random effects from the model lmer(state curiosity ~evidence quality + TOT + familiarity + (1 | statement) + (evidence quality | ID), data, REML = T), SR as baseline for evidence quality

Predictors	*b*	*SE*	*t*	*p*
Intercept	0.29	0.08	3.70	**<.001**
Evidence quality (SI)	-0.52	0.06	-8.53	**<.001**
Evidence quality (NR)	-0.79	0.08	-9.84	**<.001**
TOT	-0.05	0.01	-3.38	**.001**
Familiarity	0.03	0.01	2.19	**.029**
Random effects				
σ^2	0.58			
τ_{00} Statement	0.08			
τ_{00} ID	0.28			
τ_{11} ID. evidence quality SI	0.06			
τ_{11} ID. evidence quality NR	0.23			
ρ_{01} ID. evidence quality SI	-0.45			
ρ_{01} ID. evidence quality NR	-0.57			
ICC	0.36			
N Statement	270			
N ID	59			
Observations	5024			
Marginal R^2	0.111			
Conditional R^2	0.427			

Table A2
Usefulness: Fixed and random effects from the model lmer(usefulness ~evidence quality + TOT + familiarity + (1 | statement) + (evidence quality | ID), data, REML = T), SR as baseline for evidence quality

Predictors	*b*	*SE*	*t*	*p*
Intercept	0.68	0.05	13.52	**<.001**
Evidence quality (SI)	-1.23	0.07	-16.99	**<.001**
Evidence quality (NR)	-1.33	0.07	-18.19	**<.001**
TOT	-0.01	0.01	-0.66	.510
Familiarity	0.02	0.01	1.29	.197
Random effects				
σ^2	0.37			
τ_{00} Statement	0.10			
τ_{00} ID	0.07			
τ_{11} ID. evidence quality SI	0.15			
τ_{11} ID. evidence quality NR	0.15			
ρ_{01} ID. evidence quality SI	-0.37			
ρ_{01} ID. evidence quality NR	-0.41			
ICC	0.36			
N Statement	270			
N ID	59			
Observations	5024			
Marginal R^2	0.403			
Conditional R^2	0.619			

Table A3
Confidence: Fixed and random effects from the model lmer(confidence ~evidence quality + TOT + familiarity + (1 | statement) + (evidence quality | ID), data, REML = T), SI as baseline for evidence quality

Predictors	*b*	*SE*	*t*	*p*
Intercept	-0.12	0.07	-1.73	.084
Evidence quality (SR)	0.20	0.07	3.04	**.002**
Evidence quality (NR)	0.21	0.07	3.00	**.003**
TOT	-0.01	0.01	-1.01	.315
Familiarity	0.05	0.02	3.33	**.001**
Random effects				
σ^2	0.67			
τ_{00} Statement	0.08			
τ_{00} ID	0.21			
τ_{11} ID. evidence quality SR	0.09			
τ_{11} ID. evidence quality NR	0.12			
ρ_{01} ID. evidence quality SR	-0.46			
ρ_{01} ID. evidence quality NR	-0.21			
ICC	0.31			
N Statement	270			
N ID	59			
Observations	5024			
Marginal R^2	0.012			
Conditional R^2	0.314			

Table A4
Clarity: Fixed and random effects from the model lmer(clarity ~ evidence quality + TOT +familiarity + (1 | statement) + (evidence quality | ID), data, REML = T), NR as baseline for evidence quality

Predictors	*b*	*SE*	*t*	*p*
Intercept	0.23	0.08	2.99	**.003**
Evidence quality (SR)	-0.18	0.07	-2.66	**.008**
Evidence quality (SI)	-0.25	0.06	-3.93	**<.001**
TOT	0.06	0.01	4.46	**<.001**
Familiarity	0.04	0.01	2.74	**.006**
Random effects				
σ^2	0.60			
τ_{00} Statement	0.10			
τ_{00} ID	0.26			
τ_{11} ID. evidence quality SR	0.09			
τ_{11} ID. evidence quality SI	0.07			
ρ_{01} ID. evidence quality SR	-0.34			
ρ_{01} ID. evidence quality SI	-0.26			
ICC	0.37			
N Statement	270			
N ID	59			
Observations	5024			
Marginal R^2	0.016			
Conditional R^2	0.378			

Table A5
Number of rounds (statements read): Fixed and random effects from the model glmer(rounds ~ state curiosity+ familiarity + (1 | question) + (1 | ID), family=poisson, data, REML = T)

Predictors	*b*	*SE*	*z*	*p*
Intercept	1.46	0.07	22.34	**<.001**
State curiosity	-0.05	0.03	-1.77	.077
Familiarity	-0.04	0.02	-1.88	.061
Random effects				
σ^2	0.19			
τ_{00} ID	0.21			
τ_{00} Question	0.01			
ICC	0.54			
N Questions	18			
N ID	59			
Observations	1036			
Marginal R^2	0.007			
Conditional R^2	0.543			

Table A6
Proportion of statements read: Fixed and random effects from the model lmer(nFreq ~ evidence quality + state curiosity+ (1 | ID), data, REML =

T), SR as baseline for evidence quality.

Predictors		*b*	*SE*	*t*	*p*
Intercept		0.36	0.02	14.61	**<.001**
Evidence quality (SI)		-0.06	0.02	-2.75	**.007**
Evidence quality (NR)		-0.07	0.03	-2.70	**.008**
State curiosity		0.05	0.02	1.90	.060
Random effects					
σ^2	0.01				
$\tau_{00\ ID}$	0.02				
ICC	0.71				
N_{ID}	59				
Observations	177				
Marginal R^2	0.071				
Conditional R^2	0.733				

Table A7
Exploration: Fixed and random effects from the model glmer(exploration ~ evidence quality + state curiosity + TOT + (evidence quality | ID), family=binomial, data, REML = T), SR as baseline for evidence quality

Predictors		*b*	*SE*	z	*p*
Intercept		1.75	0.29	6.02	**<.001**
Evidence quality (SI)		0.34	0.22	1.55	.121
Evidence quality (NR)		0.70	0.28	2.49	**.013**
State curiosity		0.03	0.05	0.52	.602
TOT		-0.42	0.05	-8.55	**<.001**
Random effects					
σ^2	3.29				
$\tau_{00\ ID}$	3.53				
$\tau_{11\ ID.\ evidence\ quality\ SI}$	0.62				
$\tau_{11\ ID.\ evidence\ quality\ NR}$	1.33				
$\rho_{01\ ID.\ evidence\ quality\ SI}$	-0.47				
$\rho_{01\ ID.\ evidence\ quality\ NR}$	-0.34				
ICC	0.50				
N_{ID}	59				
Observations	3994				
Marginal R^2	0.041				
Conditional R^2	0.518				

Table A8
Accuracy: Fixed and random effects from the model glmer(accuracy ~ state curiosity+ familiarity + rounds + (1 | question) + (1 | ID), family=binomial, data, REML = T).

Predictors		*b*	*SE*	z	*p*
Intercept		0.65	0.25	2.59	**.010**
State curiosity		0.17	0.13	1.35	.178
Familiarity		0.06	0.10	0.61	.539
Rounds		0.12	0.03	3.71	**<.001**
Random effects					
σ^2	3.29				
$\tau_{00\ ID}$	0.34				
$\tau_{00\ Question}$	0.51				
ICC	0.20				
$N_{Question}$	18				
N_{ID}	59				
Observations	1036				
Marginal R^2	0.039				
Conditional R^2	0.236				

Table A9
Gaze duration: Fixed and random effects from the model lmer(gaze ~ evidence quality * state curiosity + word frequency + word length + (1 | word) + (1 | statement) + (1 | ID), data, REML = T), SR as baseline for evidence quality.

Predictors		*b*	*SE*	*t*	*p*
Intercept		5.45	0.02	231.98	**<.001**
Evidence quality (SI)		-0.01	0.01	-1.00	.317
Evidence quality (NR)		-0.02	0.01	-1.77	.076
State curiosity		-0.00	0.00	-1.19	.234
Word frequency		-0.06	0.01	-8.87	**<.001**
Word length		0.16	0.01	26.58	**<.001**
Evidence quality (SI)*Curiosity		0.01	0.01	1.83	.067
Evidence quality (NR) *Curiosity		0.01	0.01	1.40	.160
Random effects					
σ^2	0.18				
$\tau_{00\ Word}$	0.01				
$\tau_{00\ Statement}$	0.00				
$\tau_{00\ ID}$	0.03				
ICC	0.19				
N_{Word}	1636				
$N_{Statement}$	270				
N_{ID}	56				
Observations	36505				
Marginal R^2	0.144				
Conditional R^2	0.307				

(continued on next column)

Table A9 (*continued*)

Table A10
Go-past time: Fixed and random effects from the model lmer(gpt ~ evidence quality * state curiosity + word frequency + word length + (1 | word) + (1 | statement) + (evidence quality | ID), data, REML = T), SR as baseline for evidence quality.

Predictors		*b*	*SE*	*t*	*p*
Intercept		5.98	0.03	177.71	**<.001**
Evidence quality (SI)		-0.04	0.02	-1.93	.053
Evidence quality (NR)		-0.01	0.02	-0.37	.707
State curiosity		0.01	0.01	0.73	.464
Word frequency		0.03	0.03	1.18	.239
Word length		0.34	0.03	12.80	**<.001**
Evidence quality (SI)*Curiosity		0.02	0.01	1.79	.074
Evidence quality (NR) *Curiosity		0.02	0.01	1.56	.119
Random effects					
σ^2	0.47				
$\tau_{00\ Word}$	0.42				
$\tau_{00\ Statement}$	0.01				
$\tau_{00\ ID}$	0.04				
$\tau_{11\ ID.\ evidence\ quality\ SI}$	0.00				
$\tau_{11\ ID.\ evidence\ quality\ NR}$	0.01				
$\rho_{01\ ID.\ evidence\ quality\ SI}$	-0.41				
$\rho_{01\ ID.\ evidence\ quality\ NR}$	-0.22				
ICC	0.50				
N_{Word}	1636				
$N_{Statement}$	270				
N_{ID}	56				
Observations	36505				
Marginal R^2	0.084				
Conditional R^2	0.539				

Table A11
Skipping rate: Fixed and random effects from the model glmer(skip ~ evidence quality * state curiosity + word frequency + word length + (1 | word) + (1 | statement) + (evidence quality | ID), data, family=binomial, REML = T), SR as baseline for evidence quality

Predictors		*b*	*SE*	z	*p*
Intercept		-1.05	0.09	-11.13	**<.001**
Evidence quality (SI)		-0.23	0.08	-3.05	**.002**
Evidence quality (NR)		-0.26	0.08	-3.43	**.001**
State curiosity		-0.05	0.02	-2.38	**.018**
Word frequency		0.11	0.03	3.37	**.001**
Word length		-0.66	0.03	-20.05	**<.001**
Evidence quality (SI)*Curiosity		-0.02	0.03	-0.59	.559
Evidence quality (NR) *Curiosity		0.07	0.03	1.93	.054
Random effects					
σ^2	3.29				

(*continued on next page*)

Table A11 (*continued*)

Predictors	*b*	*SE*	*z*	*p*
τ_{00} Word	0.28			
τ_{00} Statement	0.14			
τ_{00} ID	0.36			
τ_{11} ID. evidence quality SI	0.08			
τ_{11} ID. evidence quality NR	0.07			
ρ_{01} ID. evidence quality SI	-0.10			
ρ_{01} ID. evidence quality NR	-0.00			
ICC	0.20			
N Word	1636			
N Statement	270			
N ID	56			
Observations	53792			
Marginal R^2	0.122			
Conditional R^2	0.296			

Table A12
Total fixation duration: Fixed and random effects from the model lmer (tfd ~ evidence quality * state curiosity + word frequency + word length + (1 | word) + (1 | statement) + (evidence quality | ID), data, REML = T), SR as baseline for evidence quality.

Predictors	*b*	*SE*	*t*	*p*
Intercept	6.05	0.03	175.85	**<.001**
Evidence quality (SI)	-0.01	0.03	-0.42	.678
Evidence quality (NR)	-0.08	0.03	-2.93	**.003**
State curiosity	0.02	0.01	3.32	**.001**
Word frequency	-0.10	0.01	-10.21	**<.001**
Word length	0.22	0.01	22.54	**<.001**
Evidence quality (SI)*Curiosity	0.03	0.01	2.81	**.005**
Evidence quality (NR) *Curiosity	0.04	0.01	4.31	**<.001**
Random effects				
σ^2	0.37			
τ_{00} Word	0.03			
τ_{00} Statement	0.02			
τ_{00} ID	0.05			
τ_{11} ID. evidence quality SI	0.01			
τ_{11} ID. evidence quality NR	0.01			
ρ_{01} ID. evidence quality SI	0.02			
ρ_{01} ID. evidence quality NR	0.13			
ICC	0.23			
N Word	1636			
N Statement	270			
N ID	56			
Observations	36505			
Marginal R^2	0.148			
Conditional R^2	0.343			

Table A13
Delayed response consistency (i.e., choosing the same response to the questions in the lab task and in the online survey) as an indicator of memory for the topics: Fixed and random effects from the model glmer (same ~ state curiosity + familiarity + rounds + Delay + (1 | Question) + (1 | ID), data, family=binomial, REML = T).

Predictors	*b*	*SE*	z	*p*
Intercept	1.17	0.18	6.42	**<.001**
Mean state curiosity	0.13	0.13	1.00	.319
Familiarity	0.15	0.11	1.37	.171
Rounds	0.31	0.10	3.13	**.002**
Delay	-0.03	0.09	-0.36	.723
Random effects				
σ^2	3.29			
τ_{00} ID	0.09			
τ_{00} Question	0.43			
ICC	0.14			
N Question	18			
N ID	52			
Observations	919			
Marginal R^2	0.032			
Conditional R^2	0.166			

Table A14
State curiosity explained by trait curiosity: Fixed and random effects from the model lmer(state curiosity ~ Joyous Exploration + Deprivation Sensitivity + Stress Tolerance + Thrill Seeking + Overt Social Curiosity + Covert Social Curiosity + (1 | ID), data, REML = T).

Predictors	*b*	*SE*	*t*	*p*
Intercept	-0.03	0.05	-0.50	.616
Joyous Exploration	0.08	0.07	1.15	.250
Deprivation Sensitivity	0.03	0.06	0.46	.646
Stress Tolerance	-0.01	0.06	-0.14	.886
Thrill Seeking	-0.19	0.06	-3.04	**.002**
Overt Social Curiosity	0.02	0.06	0.26	.796
Covert Social Curiosity	-0.02	0.06	-0.31	.754
Random effects				
σ^2	0.25			
τ_{00} ID	0.14			
ICC	0.36			
N ID	52			
Observations	919			
Marginal R^2	0.080			
Conditional R^2	0.413			

Table A15
Number of rounds (statements read) by trait curiosity: Fixed and random effects from the model glmer(rounds ~ Joyous Exploration + Deprivation Sensitivity + Stress Tolerance + Thrill Seeking + Overt Social Curiosity + Covert Social Curiosity + (1 | ID), data, family=poisson, REML = T)

Predictors	*b*	*SE*	*z*	*p*
Intercept	1.47	0.06	24.02	<**.001**
Joyous Exploration	0.01	0.07	0.20	.844
Deprivation Sensitivity	-0.03	0.07	-0.44	.662
Stress Tolerance	0.00	0.07	0.04	.966
Thrill Seeking	-0.09	0.07	-1.25	.210
Overt Social Curiosity	0.12	0.07	1.82	.069
Covert Social Curiosity	-0.07	0.06	-1.04	.297
Random effects				
σ^2	0.19			
τ_{00} ID	0.18			
ICC	0.49			
N ID	52			
Observations	919			
Marginal R^2	0.066			
Conditional R^2	0.525			

Table A16
Exploration by trait curiosity: Fixed and random effects from the model glmer(exploration ~ Joyous Exploration + Deprivation Sensitivity + Stress Tolerance + Thrill Seeking + Overt Social Curiosity + Covert Social Curiosity + (1 | ID), data, family= binomial, REML = T).

Predictors	*b*	*SE*	*z*	*p*
Intercept	0.87	0.06	13.57	<**.001**
Joyous Exploration	0.06	0.08	0.76	.449
Deprivation Sensitivity	-0.02	0.08	-0.20	.840
Stress Tolerance	0.02	0.07	0.29	.774
Thrill Seeking	-0.09	0.07	-1.19	.236
Overt Social Curiosity	0.02	0.07	0.22	.823
Covert Social Curiosity	-0.10	0.07	-1.49	.135
Random effects				
σ^2	0.32			
τ_{00} ID	0.19			
ICC	0.37			
N ID	52			
Observations	919			
Marginal R^2	0.035			
Conditional R^2	0.391			

Table A17

Thinking about the topics: Fixed and random effects from the model glmer(think ~ state curiosity + familiarity + rounds + Joyous Exploration + Deprivation Sensitivity + Stress Tolerance + Thrill Seeking + Overt Social Curiosity + Covert Social Curiosity + (1 | ID), data, family=binomial, REML = T).

Predictors		*b*	*SE*	*z*	*p*
Intercept		-8.96	3.15	-2.85	**.004**
Mean state curiosity		0.69	2.28	0.30	.762
Familiarity		-0.03	1.22	-0.02	.981
Rounds		0.46	2.05	0.23	.821
Joyous Exploration		11.72	5.21	2.25	**.024**
Deprivation Sensitivity		15.91	4.24	3.75	**<.001**
Stress Tolerance		6.92	4.14	1.67	.094
Thrill Seeking		-15.20	4.21	-3.61	**<.001**
Overt Social Curiosity		-4.09	4.19	-0.98	.329
Covert Social Curiosity		3.25	2.23	1.46	.144
Random effects					
σ^2	3.29				
$\tau_{00\ ID}$	12910.70				
ICC	1.00				
N_{ID}	52				
Observations	919				
Marginal R^2	0.049				
Conditional R^2	1.000				

Table A18
Searching for information about the topics: Fixed and random effects from the model glmer(search ~ state curiosity + familiarity + rounds + Joyous Exploration + Deprivation Sensitivity + Stress Tolerance + Thrill Seeking + Overt Social Curiosity + Covert Social Curiosity + (1 | ID), data, family=binomial, REML = T).

Predictors		*b*	*SE*	*z*	*p*
Intercept		-17.09	4.47	-3.82	**<.001**
Mean state curiosity		0.16	3.34	0.05	.963
Familiarity		0.04	1.79	0.02	.984
Rounds		0.42	1.86	0.23	.821
Joyous Exploration		-0.22	3.26	-0.07	.946
Deprivation Sensitivity		0.80	3.32	0.24	.811
Stress Tolerance		-0.65	3.51	-0.19	.852
Thrill Seeking		-1.39	4.74	-0.29	.769
Overt Social Curiosity		0.31	3.69	0.09	.932
Covert Social Curiosity		-0.25	2.65	-0.10	.924
Random effects					
σ^2	3.29				
$\tau_{00\ ID}$	7224.57				
ICC	1.00				
N_{ID}	52				
Observations	919				
Marginal R^2	0.00				
Conditional R^2	1.000				

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2026.106598.

Data availability

Data and materials for experiments are available at an Open Science Framework (OSF) repository (https://osf.io/2zcu6/, Simola, Julku, Dawson, & Österman, 2026). Supplementary material, analysis code, and data underlying statistics, figures, and main conclusions are available at an OSF repository (https://osf.io/pcq3u/, Julku, Simola, Dawson, & Österman, 2026). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon reasonable request.

References

Averbeck, B. B. (2015). Theory of choice in bandit, information sampling and foraging tasks. *PLoS Computational Biology, 11*(3), Article e1004164. https://doi.org/10.1371/journal.pcbi.1004164

Baranes, A., Oudeyer, P.-Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision Research, 117*, 81–90. https://doi.org/10.1016/j.visres.2015.10.009

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berlyne, D. E. (1954a). An experimental study of human curiosity. *British Journal of Psychology, 45*(4), 256–265.

Berlyne, D. E. (1954b). A theory of human curiosity. *British Journal of Psychology, 45*(3), 180–191.

van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS One, 11*(10), Article e0165274. https://doi.org/10.1371/journal.pone.0165274

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Dan, O., Leshkowitz, M., Livnat, O., & Hassin, R. R. (2025). Urges now, interests later: On the factors and dynamics of epistemic curiosity. *Cognition, 259*, Article 106107. https://doi.org/10.1016/j.cognition.2025.106107

Dawson, C., Julku, H., Pihlajamäki, M., Kaakinen, J. K., Schooler, J. W., & Simola, J. (2024). Evidence-based scientific thinking and decision-making in everyday life. *Cognitive Research: Principles and Implications, 9*, 50. https://doi.org/10.1186/s41235-024-00578-2

Dubey, R., Griffiths, T. L., & Lombrozo, T. (2022). If it's important, then I'm curious: Increasing perceived usefulness stimulates curiosity. *Cognition, 226*, Article 105193. https://doi.org/10.1016/j.cognition.2022.105193

Eschmann, K. C. J., Pereira, D. F. M. M., Valji, A., Dehmelt, V., & Gruber, M. J. (2023). Curiosity and mesolimbic functional connectivity drive information seeking in real life. *Social Cognitive and Affective Neuroscience, 18*(1), 1–9. https://doi.org/10.1093/scan/nsac050

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of Experimental Psychology: Human Perception and Performance, 46*(10), 1201–1221. https://doi.org/10.1037/xhp0000743

Fayn, K., Silvia, P. J., Dejonckheere, E., Verdonck, S., & Kuppens, P. (2019). Confused or curious? Openness/intellect predicts more positive interest-confusion relations. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences, 117*(5), 1016–1033. https://doi.org/10.1037/pspp0000257

FitzGibbon, L., Lau, J. K. L., & Murayama, K. (2020). The seductive lure of curiosity: information as a motivationally salient reward. *Current Opinion in Behavioral Sciences, 35*, 21–27. https://doi.org/10.1016/j.cobeha.2020.05.014

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology, 97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality, 56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B: Methodological, 41*(2), 148–177.

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience, 19*(12), 758–770. https://doi.org/10.1038/s41583-018-0078-0

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences, 17*(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Gross, M. E., Araujo, D. B., Zedelius, C. M., & Schooler, J. W. (2019). Is perception the missing link between creativity, curiosity and schizotypy? Evidence from spontaneous eye-movements and responses to auditory oddball stimuli. *NeuroImage, 202*, Article 116125. https://doi.org/10.1016/j.neuroimage.2019.116125

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron, 84*, 486–496. https://doi.org/10.1016/j.neuron.2014.08.060

Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in Cognitive Sciences, 23*(12), 1014–1025. https://doi.org/10.1016/j.tics.2019.10.003

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hidi, S. E., & Renninger, K. A. (2019). Interest development and its relation to curiosity: Needed neuroscientific research. *Educational Psychology Review, 31*(4), 833–852. https://doi.org/10.1007/s10648-019-09491-3

Hoeken, H., Timmers, R., & Schellens, P. J. (2012). Arguing about desirable consequences: What constitutes a convincing argument? *Thinking & Reasoning, 18*(3), 394–416. https://doi.org/10.1080/13546783.2012.669986

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience, 12*, 105. https://doi.org/10.3389/fnhum.2018.00105

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., Kompier, M. A. J., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology: Human Perception and Performance, 42*(6), 878–889. https://doi.org/10.1037/xhp0000189

Jach, H. K., Cools, R., Frisvold, A., Grubb, M. A., Hartley, C. A., Hartmann, J., … Gottlieb, J. (2024). Individual differences in information demand have a low dimensional structure predicted by some curiosity traits. *Proceedings of the National Academy of Sciences, 121*(45), Article e2415236121. https://doi.org/10.1073/pnas.2415236121

Jach, H. K., DeYoung, C. G., & Smillie, L. D. (2022). Why do people seek information? The role of personality traits and situation perception. *Journal of Experimental Psychology: General, 151*(4), 934–959. https://doi.org/10.1037/xge0001109

Julku, H., Simola, J., Dawson, C., & Österman, S. (2026). Eye movements during statement reading. Open Science Framework (OSF). https://osf.io/pcq3u/ [dataset underlying statistics and figures; Supplementary material].

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *The Journal of Artificial Intelligence Research, 4*, 237–285.

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. *Advances in Political Psychology, 38*, 179–199. https://doi.org/10.1111/pops.12396

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963–973.

Kashdan, T. B., Disabato, D. J., Goodman, F. R., & McKnight, P. E. (2020). The five-dimensional curiosity scale revised (5DCR): Briefer subscales while separating overt and covert social curiosity. *Personality and Individual Differences, 157*, Article 109836. https://doi.org/10.1016/j.paid.2020.109836

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*, 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7*(5), Article e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development, 85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Knuuti, J. (2020). *Kauppatavarana terveys - selviydy terveysväitteiden viidakossa*. Minerva Kustannus Oy.

Kobayashi, K., & Kable, J. W. (2024). Neural mechanisms of information seeking. *Neuron, 112*(11), 1741–1756. https://doi.org/10.1016/j.neuron.2024.04.008

Ligneul, R., Mermillod, M., & Morisseau, T. (2018). From relief to surprise: Dual control of epistemic curiosity in the human brain. *NeuroImage, 181*, 490–500. https://doi.org/10.1016/j.neuroimage.2018.07.038

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences, 44*, 1585–1595. https://doi.org/10.1016/j.paid.2008.01.014

Litman, J. A., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition & Emotion, 19*(4), 559–582. https://doi.org/10.1080/02699930441000427

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment, 80*(1), 75–86. https://doi.org/10.1207/S15327752JPA8001_16

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Lüdecke, D. (2025). sjPlot: Data Visualization for Statistics in Social Science. https://CRAN.R-project.org/package=sjPlot.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *The Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S., & Ginter, F. (2015). Towards Universal Web Parsebanks. In *Third International Conference on Dependency Linguistics (Depling 2015), Uppsala, Sweden*.

Lydon-Staley, D. M., Zhou, D., Blevins, A. S., Zurn, P., & Bassett, D. S. (2021). Hunters, busybodies and the knowledge network building associated with deprivation curiosity. *Nature Human Behavior, 5*(3), 327–336. https://doi.org/10.1038/s41562-020-00985-7

MacPherson, L., Magidson, J. F., Reynolds, E. K., Kahler, C. W., & Lejuez, C. W. (2010). Changes in sensation seeking and risk-taking propensity predict increases in alcohol use among early adolescents. *Alcoholism: Clinical and Experimental Research, 34*(8), 1400–1408. https://doi.org/10.1111/j.1530-0277.2010.01223.x

Markey, A., & Loewenstein, G. (2014). Curiosity. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 228–245). Routledge.

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General, 145*(3), 266–272. https://doi.org/10.1037/xge0000140

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., … Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191–215. https://doi.org/10.1037/dec0000033

Metcalfe, J., Schwartz, B. L., & Eich, T. S. (2020). Epistemic curiosity and the region of proximal learning. *Current Opinion in Behavioral Sciences, 35*, 40–47. https://doi.org/10.1016/j.cobeha.2020.06.007

Murayama, K. (2022). A Reward-learning framework of knowledge acquisition: An Integrated account of curiosity, interest, and intrinsic–extrinsic rewards. *Psychological Review, 129*(1), 175–198. https://doi.org/10.1037/rev0000349

Pekrun, R. (2019). The murky distinction between curiosity and interest: State of the art and future prospects. *Educational Psychology Review, 31*, 905–914. https://doi.org/10.1007/s10648-019-09512-1

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science, 21*(9), 1300–1310. https://doi.org/10.1177/0956797610378686

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition, 122*, 86–90. https://doi.org/10.1016/j.cognition.2011.08.014

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion : Theory, methods, research* (pp. 92–120). NY: Oxford University Press.

Schumacher, A., Spitzer, M. W. H., Jach, H., Kammerer, Y., Scharinger, C., & Bardach, L. (2026). From clicks to curiosity: Exploring self-directed information seeking as a behavioral manifestation of curiosity. *Journal of Experimental Psychology. Applied*.

Silvia, P. J. (2008). Appraisal components and emotion traits: Examining the appraisal basis of trait curiosity. *Cognition and Emotion, 22*(1), 94–113. https://doi.org/10.1080/02699930701298481

Silvia, P. J. (2010). Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts, 4*(2), 75–80. https://doi.org/10.1037/a0017081

Simola, J., Julku, H., Dawson, C., & Österman, S. (2026). Neurocognitive Signatures of Scientific Evidence-based Decision Making. Open Science Framework (OSF). https://osf.io/2zcu6/ [dataset for task codes and stimulus materials].

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*, 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Speekenbrink, M. (2022). Chasing unknown bandits: Uncertainty guidance in learning and decision making. *Current Directions in Psychological Science, 31*(5), 419–427. https://doi.org/10.1177/09637214221105051

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology, 53*, 168–179. https://doi.org/10.1016/j.jmp.2008.11.002

Stojic, H., Schulz, E., Analytis, P. P., & Speekenbrink, M. (2020). It's new, but is it good? How generalization and uncertainty guide the exploration of novel options. *Journal of Experimental Psychology: General, 149*(10), 1878. https://doi.org/10.1037/xge0000749

Ten, A., Kaushik, P., Oudeyer, P.-Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications, 12*(1), 5972. https://doi.org/10.1038/s41467-021-26196-w

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 832–842. https://doi.org/10.1037/a0029669

Vogl, E., Pekrun, R., Murayama, K., & Loderer, K. (2020). Surprised–curious–confused: Epistemic emotions and knowledge exploration. *Emotion, 20*(4), 625. https://doi.org/10.1037/emo0000578

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General, 143*(6), 2074. https://doi.org/10.1037/a0038199

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behavior, 2*, 915–924. https://doi.org/10.1038/s41562-018-0467-4